



# The awareness assessment model: measuring awareness and collaboration support over participant's perspective

Márcio José Mantau<sup>1</sup> · Fabiane Barreto Vavassori Benitti<sup>2</sup>

Accepted: 19 February 2024

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

## Abstract

Awareness has been a valuable concept in collaborative systems since its formation, an essential part of groupware. Awareness research followed the evolution of the whole field over the last decades. We can see the progress in a mutual understanding of awareness and developing concepts and technology of awareness support. An efficient awareness mechanism ensures a better understanding and, consequently, a better projection of future actions; in contrast, the lack of these mechanisms undermines comprehension and prevents participants from projecting their work accordingly. Few works present methods or processes that assist in providing awareness in groupware systems; most common strategies focus on the design/development stages or are ad-hoc evaluation models. Furthermore, there are no standardized tests for awareness assessment; thus, measures must be established to assess awareness and identify the criteria for achieving awareness indicators. This work establishes an assessment model for collaborative interfaces by analyzing the awareness mechanisms provided from the participant's viewpoint. In this model, we consider the participant's skill in understanding the awareness and the difficulty involved, providing advances toward designing, developing, and evaluating groupware systems. The proposed assessment model allows us to measure the awareness support provided considering the collaboration, workspace, and contextual awareness perspectives. Assuming a plural collaborative environment, where different participants with different skills, knowledge, and wisdom meet and interact, this model seeks to build a more faithful representation of these existing profiles across a broad spectrum of individual abilities.

**Keywords** Awareness · Collaboration · Communication · Cooperation · Coordination · 3C model

## 1 Introduction

Connecting people, enabling individuals to collaborate, and supporting social interaction are parts of the human essence, and collaborative systems, also known as groupware, are great environments to provide them. Applications such as

blogs, wiki pages, instant messaging, digital voicemail, videoconferencing, chat forums, social media, online gaming, etc., are daily tools that allow people to connect, interact, and work together.

Ellis et al. [30] defined groupware as a computer-based system that supports two or more people engaged in a common task or objective and provides an interface for a shared environment. Groupware must provide support for three fundamental pillars: communication, coordination, and cooperation (3C collaboration model) [40], and, to support each of these collaboration pillars, groupware must make available cues/information allowing participants to communicate, coordinate, and cooperate. This support involves a fundamental element of a collaborative system: the awareness [27].

Awareness has been a significant concept in the Computer Supported Cooperative Work (CSCW) field since its formation [95] and is an essential part of groupware [39, 67]. This fundamental concept can be defined as “the understanding

---

Márcio José Mantau and Fabiane Barreto Vavassori Benitti contributed equally to this work.

✉ Márcio José Mantau  
marcio.mantau@gmail.com

✉ Fabiane Barreto Vavassori Benitti  
fabiane.benitti@ufsc.br

<sup>1</sup> Software Engineering Department, Universidade do Estado de Santa Catarina, Rua Dr. Getúlio Vargas, 2022, Ibirama, Santa Catarina 89140-000, Brazil

<sup>2</sup> Department of Informatics and Statistics, Universidade Federal de Santa Catarina, Carvoeira, Florianópolis, Santa Catarina 88040-370, Brazil

of the activities of others involved and which provides a context for their activities” [28, 88] or “set of processes in which participants recognize, organize and make sense of the stimuli received from the environment that they are in” [88, 89]. In this context, awareness ensures that individual activities are tuned with the group’s objectives and enables collaboration. Without awareness, there is no possibility of collective work, and the group will be just an incoherent set of isolated pieces [16]. The awareness of individual and group tasks is critical to the success of the collaboration process [27].

We consider awareness the backbone of a collaborative environment; all collaborative concepts are archived through it. In this sense, we define awareness as a process that occurs at three basic levels of abstraction: representation, understanding, and projection. An efficient awareness mechanism ensures a better experience and, consequently, a better projection of future actions. In contrast, the lack of awareness mechanisms undermines comprehension and prevents participants from projecting their work accordingly.

Awareness elements help people move between individual and shared activities, provide a context for interpreting others, allow anticipation of actions, and reduce the effort expended to coordinate activities [38]. Awareness enables a given user to perceive the sensation of working in a group [19]. Information awareness is essential for coordination, cooperation, and communication (3C collaboration model). It allows for building a shared understanding of the task, being aware of the activities of other participants, knowing the progress of your work and that of the whole group, and transmitting group strategies and plans [36].

Over the last three decades, different awareness types have emerged in the literature. In a non-exhaustive list of the main awareness types, we quickly identify task awareness; concept awareness [42]; context awareness [37]; workspace awareness [38, 43]; historical awareness [56]; social awareness [15]; presence awareness; group-structural awareness [29]; group awareness [52, 54]; situation awareness [3]; behavioral awareness [14]; cognitive awareness [51]; knowledge awareness [31, 105]; user awareness [47, 98]; activity awareness [74, 77]; and, we-awareness [95].

Certain works [5, 35, 63, 85] present a broader list of awareness elements. Detailed background about awareness origins, the early ethnographic, and the technology studies that brought up fundamental insights can be found in [39].

## 1.1 The awareness problem

Awareness is a well-known concern but still not fully reached concept in collaborative environments. According to Gross [39], the awareness concept remains difficult to grasp, and future research should achieve a better understanding of supporting awareness and effortless coordination

and conceiving and testing novel technology that guarantees awareness aspects and minimum coordination effort. Moreover, since the 1980 s, there has been no consensus about the awareness issue and how to understand it [83].

Awareness is a multi-factorial problem, and few papers are addressing it from a broad point of view. Finding a good starting point in the literature can be challenging for novice awareness designers [67] because they must reinvent awareness from their own experience of what it is, how it works, and how it is used [19].

The lack of awareness in collaborative interfaces can be harmful, and some problems must be considered when using awareness elements like information overload, intrusiveness, and privacy [46, 82]. At a high level, both people may differ in their understandings, and the individual’s awareness may change as their background and received stimuli change. People have different capabilities in representing, understanding, and projecting human actions through interface [63]. Furthermore, considering awareness as an understanding or even a mental state of an individual about a specific object or environmental stimulus, we believe that the design, development, and evaluation approaches should consider awareness from the participant’s perspective.

These problems raise some questions that should be considered when developing collaborative applications [38, 62, 63]: (1) How can we present the awareness information in the interface? (2) Which information is relevant and which can be ignored? (3) What information can be shared? (4) Who can have access to this information? However, it is difficult to find methods that allow designers to develop collaborative applications centered on awareness aspects [19]. Awareness is a concept that promises to improve the usability of collaborative applications [93]; however, no clear overall picture of awareness has yet emerged from a collaborative perspective [58].

## 1.2 The assessment problem

The evaluation of collaborative systems is more complex and challenging than conventional ones.

First, providing awareness and 3C model aspects involves dealing with two main trade-offs: (1) informativeness versus privacy: if the current status of a person is visible enough to be helpful to others, it often violates that person’s privacy [79]; and (2) information versus overloading: the lack of awareness support may compromise the group’s activities; conversely, it is essential to avoid information overload, presenting just relevant information [62].

Second, collaborative evaluation belongs to more than one temporal dimension. It is complex to obtain data about each in just one way [4]: (1) individual information is gathered focusing on events occurring in a time frame of a few minutes or even seconds; (2) group information is gathered

addressing activities occurring in the range of several minutes and hours; and (3) organizational information concerns much longer time frames, usually in the order of days, months, and even years.

Third, research fails to address conceptual frameworks covering the four trends: theoretical frameworks, context modeling, collaborative design, and awareness [13]. It remains necessary to establish a theoretical framework for analyzing or modeling cooperative work and specifying requirements of computer-based systems to support cooperative work [21]. A practical, holistic framework may conduct organizations and other social entities in their effort to design, evaluate, and acquire collaboration systems that can support their needs [21]. It is hard to generate adaptation rules automatically, and no frameworks help designers to incorporate semi-automatically users' feedback [2].

Fourth, few works present methods or processes that assist in providing aspects of awareness in groupware systems [19], and there are no standardized tests for awareness assessment [75]. There is a need to establish measures to assess awareness [75] and identify the criteria for achieving awareness indicators [67]. Therefore, research toward awareness assessment strategies is necessary to measure the quality of awareness support provided by the collaborative applications under development and/or evolution [5].

Regarding collaborative assessment strategies, Lopez and Guerrero [58] identified through a systematic literature review that almost half of 83 selected papers (42%) did not specify the approach used in the groupware assessment, and just 6% presented an example proposal. In a recent mapping study [59], it was identified that 45% of papers with a substantial problem in data analysis and 9% just presented a solution proposal paper, clearly demonstrating the need for greater rigor in adopting methods, procedures, and materials when conducting groupware evaluation.

### 1.3 Aims of the study

In this work, we contribute to developing an awareness assessment model to access awareness and collaboration support through measuring awareness mechanisms from the participant's viewpoint. The model considers the participant's skill in understanding the awareness information provided by the application and the difficulty involved in perceiving each awareness piece.

We assume the following assumptions:

- (1) Awareness is an individual understanding of a particular environmental object or stimulus. It is the means available to interact with each other and involves, from the participant's viewpoint, the representation (mechanisms or elements that provide participants cues about

“what is going on”) and the understanding or consciousness of something;

- (2) Collaboration results from the participant's understanding/consciousness. The consciousness allows individuals to project their actions;
- (3) Awareness is intrinsically linked to the participant's skills in identifying, understanding, or projecting their actions. Individuals may have different awareness; likewise, the participant's understanding differs over time.

This paper is organized as follows. Section 2 presents a background of the awareness and collaboration assessment approaches described in the literature and related issues. The methodology is described in Sect. 3. The awareness assessment model is presented in section 4, detailing the assessment process and the model's conceptual view structure. Sections 5 and 6 describe strategies adopted to access the model validation through an expert panel and case study scenarios. Model reliability and dimensionality are investigated in Sects. 7 and 8. Discussions and conclusions are presented in Sects. 9 and 10. Supplementary materials are available in Appendix A.

## 2 Related work

Some ways of evaluating awareness and collaboration were presented in the recent literature [59]. The most common strategy is an ad hoc approach that involves users through experiments or case studies. Questionnaires, interviews, brainstorming, focus groups, conceptual modeling, direct observation, system logs, and static/dynamic analysis of a system were used in the assessment strategies.

Questionnaires were the main data collection tool reported [59]: by user experience [54], usability [64], NASA-TLX user workload [41], or ethnographic [47] questionnaires; by ethnographic questionnaire combined with system logs and researcher's observations [62], and system logs [63]; by participatory observations, nonstructured and mostly ad hoc interviews, and discussions [92]; by semi-structured interview and a 7-point Likert scale questionnaire combined with statistical analysis [105]; and by 7-point Likert scale ethnographic and usability questionnaire combined with researcher's observations, system logs, audio and video recordings [75].

A combination of different assessment techniques, like frameworks, guidelines, design requirements, or groupware heuristics, were also applied during development and evaluation, namely: checklist to assess awareness support in groupware systems [5, 26]; set of requirements and assessment metrics [63]; usability groupware heuristics for mobile environments [23]; frameworks or taxonomies [19, 24, 34,

67]; or questionnaires, laboratory testing, heuristic evaluation, automatic logging, and eye-tracking techniques [65].

Antunes et al. [5] developed an awareness checklist to help software designers inspect the quality of awareness support in applications under development or evolution. The checklist comprises 54 design elements and six awareness types: collaborations, Location, Context, Social, Workspace, and Situation. A recent adaptation of the checklist is presented by Do Espírito Santo et al. [26], where the authors investigate awareness support in the context of agile software engineering development.

Collazos et al. [19] elaborated a descriptive theory for groupware development to assist groupware engineers in incorporating awareness mechanisms by focusing on the aspects to be considered when designing and implementing awareness mechanisms in groupware tools.

De Souza and Barbosa [24] proposed an extension to the MoLIC (Modelling Language for Interaction as Conversation) that helps designers project collaborative applications considering the influences between users, cooperative tasks, and awareness mechanisms.

Gallardo et al. [34] proposed an awareness ontology that conceptualizes aspects relating to awareness in collaborative modeling systems. The method embraces the conceptual (steps to be carried out), methodological (aspects to be taken into account in the generation of the collaborative tool), and technological frameworks (specific IDE plug-ins to support collaborative functionality).

Molina et al. [65] proposed an evaluation approach combining subjective (e.g., the subjective perception collected by questionnaires about his/her satisfaction) and objective (eye tracking techniques) to evaluate interactive systems. This approach allows the examiner to evaluate the awareness support of collaborative systems by combining inspection (heuristic evaluation), subjective (questionnaires, interviews) and objective (automatic logging) inquiry, and usability testing lab (retrospective thinking aloud, eye tracking, and recording of the use).

Niemantsverdriet et al. [67] projected a framework for awareness designers structured into a list of design considerations to support awareness interaction that can be used during the design process.

The works of De Souza and Barbosa [24], Collazos et al. [19], and Niemantsverdriet et al. [67] have a central focus on the design and development of collaborative environments; the central focus of these works is to provide design considerations to support interaction designers during their design process.

Despite the work of Antunes et al. [5] described a checklist to inspect the quality of awareness support, this focuses mainly on the development stages and software designers. Furthermore, the validity and reliability of the proposed instrument were not verified.

The model proposed by Molina et al. [65] uses a notable set of different evaluation strategies; however, it requires a specific evaluation environment (laboratory), and the peculiarities of employing a dynamic evaluation approach, like eye-tracking or usage recording techniques, can also be a limiting aspect for its replication in scenarios with limited computational resources available to conduct experimentation. Additionally, due to the small sample size, the results should be considered preliminary.

As we can see, there remains a need to develop awareness assessment strategies for collaborative environments, aiming to evaluate the awareness support in the context of use and from the participant's perspective. There are no standardized tests to evaluate awareness [67], and it remains necessary to identify awareness evaluation criteria and establish quality indicators for collaborative environments [75].

### 3 Methods

This research was carried out in four steps. First, we perform a systematic mapping study to identify the awareness support in the collaborative system context. In the second step, we define an awareness taxonomy that contemplates a complete set of awareness elements and collaboration aspects necessary for cooperative work. In the third step, an awareness assessment model for collaborative environments is established based on systematic mapping results and awareness taxonomy. Finally, we carry out the model validation process.

#### 3.1 Awareness support identification

We performed a systematic mapping study following the guidelines presented in [55, 72, 73]. The analysis of the state-of-the-art aims to identify existing approaches, models, methodologies, or processes used in the development and evaluation of awareness and collaboration aspects in groupware systems, addressing related awareness mechanisms and/or elements to support awareness and collaboration concepts. This systematic mapping study aims to answer the following:

- *What are the approaches (e.g., models, methodologies, or processes) used in the development and evaluation of awareness and collaboration aspects in groupware systems?*
  - *Does the model consider aspects of communication, coordination, and cooperation? How are they related?*

- How was awareness related to the approach? What awareness elements were reported? How were the elements used to support it?

The key terms follow the PICO structure (population, intervention, comparison, and outcomes), as presented by [73]. We applied the logical operator OR between key terms and logical operator AND between the PICO dimensions, as exposed in Table 1.

Using IEEE, ACM, Scopus, Science Direct, Engineering Village, and Web of Science search engines, we performed the search in a four-step path:

*Step 1. Executing the query.* First, the query was adjusted according to each search engine syntax and executed by searching papers with correspondence in their title, abstract, or keywords. The obtained results were filtered by the inclusion criteria (IC):

- IC1. Papers since 2010;
- IC2. Written in English;
- IC3. Published in Journals or Proceedings.

*Step 2. Applying the exclusion criteria.* We considered papers addressing heuristic assessment methods, design guidelines, requirements or assessment approaches, or papers presenting awareness elements. Over the results obtained in the previous step, the following exclusion criteria (EC) were applied:

- EC1. Papers with restricted access to full text, short papers, posters, abstracts, or other material not peer-reviewed;
- EC2. To duplicate papers (identical ones), we considered the first result;
- EC3. To duplicate papers (extensions or a similar one), we considered the most detailed publication (more pages or most recent);
- EC4. Papers that did not address the groupware context, collaboration model, or awareness;
- EC5. Papers that did not present a groupware development or evaluation model.

The execution of the systematic mapping protocol obtained 1140 initial results. Performing steps 1 and 2 of the protocol, we selected 28 initial papers, as shown in Table 2.

*Step 3. Data extraction.* Data from selected papers is extracted using the data extraction form by using a spreadsheet to organize and document the collected data.

*Step 4. Snowballing.* Based on the results obtained in step 2, we applied the backward and forward snowballing techniques, as presented by [103]. In the forward approach, it was considered the Google Scholar indexation because it covers most of the databases used in the primary source. We performed the snowballing iteratively until no new paper was included, using both techniques in each iteration.

Over the results, we applied the ICs of step 1 and executed steps 2 and 3 (Table 3). The column BSB represents the

**Table 1** Search query

#	Concept	Keys/variations/synonyms
P	Collaborative system	Groupware OR CSCW OR “collaborative work” OR “collaborative system” OR “collaborative software”
I	Awareness	Awareness
C	Approach	Design* OR checklist OR guide* OR method* OR process OR approach OR framework OR taxonomy OR support* OR cues OR rules OR model OR “descriptive theory”
O	3C model	Communication OR cooperation OR coordination OR 3C

**Table 2** Results obtained using search engines

Step	IEEE	ACM	Scopus	Science direct	Eng. village	Web of science	Total
Initial	330	101	226	70	287	126	1140
ICs	– 0	– 0	– 4	– 4	– 0	– 0	– 8
EC1	– 82	– 43	– 10	– 0	– 85	– 5	– 228
EC2	– 52	– 0	– 167	– 31	– 83	– 80	– 413
EC3	– 1	– 0	– 2	– 0	– 0	– 0	– 3
EC4	– 149	– 25	– 32	– 24	– 44	– 35	– 309
EC5	– 44	– 27	– 7	– 8	– 59	– 6	– 151
Total	2	6	1	3	16	0	28

**Table 3** Results obtained using snowballing

Step	BSB-1	FSB-1	BSB-2	FSB-2	BSB-3	FSB-3	Total
Initial	1360	359	795	466	153	47	3180
ICs	- 1004	- 39	- 681	- 26	- 134	- 0	- 1884
EC1	- 2	- 34	- 0	- 13	- 0	- 0	- 49
EC2	- 70	- 63	- 26	- 85	- 7	- 20	- 271
EC3	- 1	- 3	- 0	- 3	- 0	- 0	- 7
EC4	- 129	- 213	- 57	- 294	- 10	- 26	- 729
EC5	- 76	- 9	- 1	- 5	- 0	- 0	- 91
Total	10	1	1	2	0	0	<b>14</b>

execution of the backward technique, and the column FSB represents the forward one. We analyzed 4320 records and selected 42 papers, considering 1140 from search engines and 3180 from snowballing. Figure 1 shows the systematic mapping results, inspired in the PRISMA flowchart [44].

The systematic mapping artifacts, including data extraction form, spreadsheets, and additional materials, are available in the dataset [60]. The full systematic mapping results are in [59].

### 3.2 Taxonomy definition method

Our Taxonomy Definition Method is based on Bailey's conceptual approach [6] and combines the guidelines presented by [66, 91, 101]. The Taxonomy Definition Method consists of four main phases, namely, Planning, Identification, Design and Construction, and Validation [91, 101], as presented in Fig. 2.

The planning phase defines the taxonomy's context and initial setting, covering the definition of meta-characteristics and the objective and subjective ending conditions. The meta-characteristics are the most comprehensive characteristics that will serve as the basis for choosing taxonomy elements [66]. The objective and subjective ending conditions are rules used to determine when to stop the interactive design and Construction process.

We collected data to define the new taxonomy in the identification phase using systematic mapping results. Then, the terms were collected, and redundancies and inconsistencies were identified and removed using a terminology control process.

The design and construction steps were performed using the phenetic analysis, classifying elements by similarity [66]. They identified different awareness characteristics or elements presented and clustered them into similar groups using classifications presented by [5, 26, 43] as a starting point. At the end of the design and construction phase, it is checked whether all objective and subjective ending conditions have been met. If so, the definition of the new taxonomy has been completed, and then the validation phase is carried out. Otherwise, a new cycle is performed.

The validation phase ensures that the taxonomy will be helpful for users to achieve their goals and strengthens their reliability and usefulness [101]. Illustrative scenarios and case studies were used to validate the taxonomy [91, 101] because they are indicated when a conceptual approach is adopted [91].

The taxonomy definition process artifacts, including raw data, spreadsheets, and validation materials, are available in the dataset [60]. Section 4.3 briefly presents the resultant awareness taxonomy incorporated into the assessment model. A full reference is available at [61].

### 3.3 Model definition method

We designed the awareness assessment process, establishing its structure, phases, activities, and work products. This awareness assessment process defines and guides the researcher through all the steps necessary to assess a collaborative environment through its support of awareness mechanisms. Section 4 describes the awareness assessment method.

Using the GQM (Goal Question Metric approach) [11, 102], the evaluation objective is defined and systematically decomposed into factors to be measured. Then, the measurement items and response format are defined, following the scale development guide [25]. These elements represent the conceptual view of the assessment model.

Using the hierarchical structure in the awareness taxonomy, the evaluation objective is defined and systematically decomposed into factors to be measured. The factors are defined to support the development of the measurement instrument (questionnaire). The response format for the measurement instrument items was based on response formats typically used in standardized questionnaires and assessment quality scales [7, 11, 25, 70].

### 3.4 Model validation

Validity constitutes a primary measurement parameter and is concerned with tests' accuracy or even measuring instrument

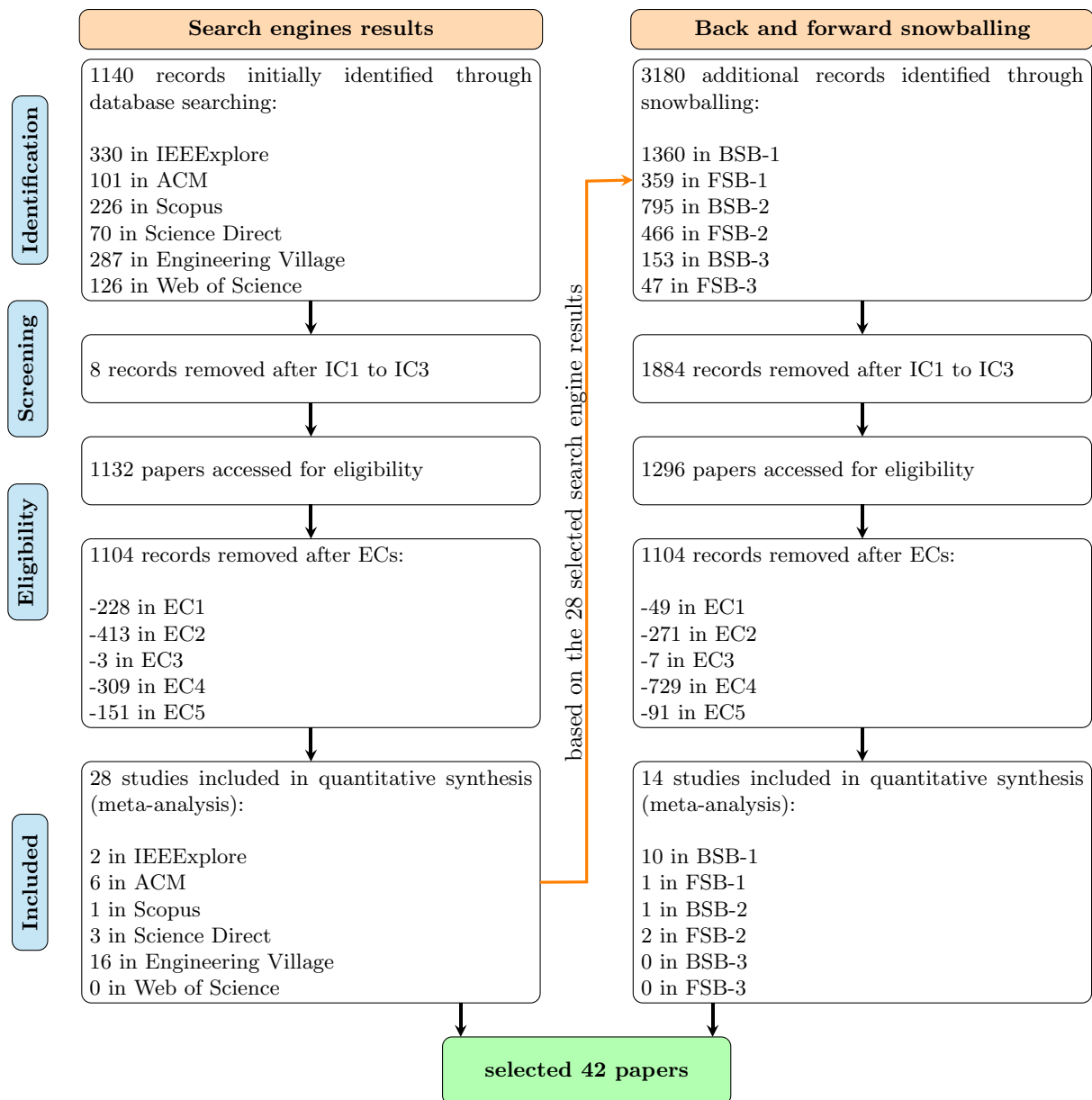


Fig. 1 Systematic mapping results (PRISMA flowchart)

calibration. Validity concerns whether the measurement is congruent with the measured property [71].

There are three types of validity: criterion validity, content, and construct validity [71, 78].

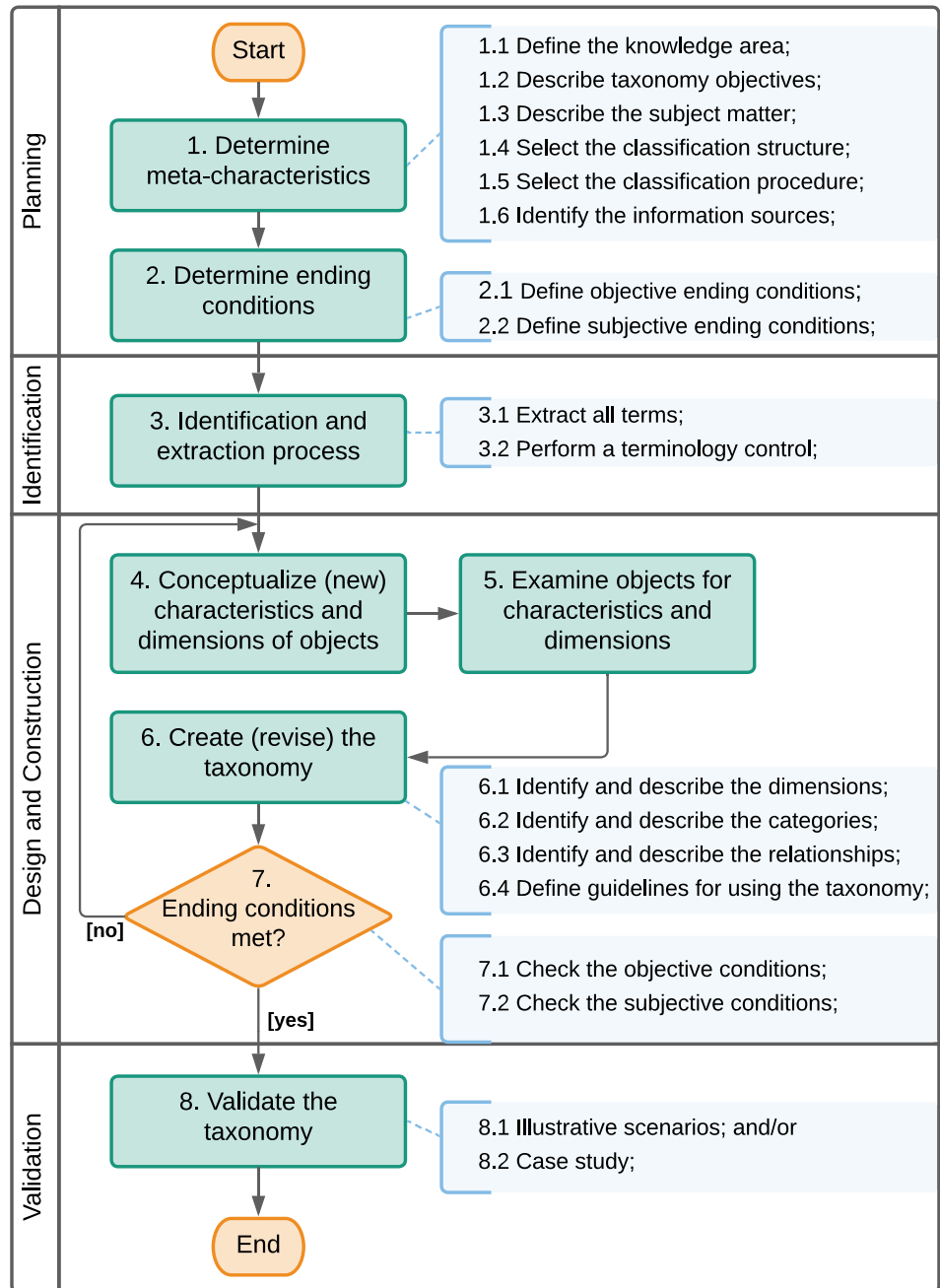
According to Richardson [78], the content of the instrument (the questions or items) are samples of different situations, and the degree to which the items represent these situations is called content validity. If a set of items constitutes a representative sample of the contents of interest, it is said to have content validity [68].

Criterion validity is characterized by the prediction about an important criterion or form observable external

to the measurement instrument itself, that is, the degree of effectiveness that a set of items has in predicting a specific performance [78].

Construct validity concerns the validation of a theory, which is reflected in a given instrument [78]. Construct validity can also be defined as the extent to which the set of items measures a theoretical latent trait [71]; or the direct way of investigating the hypothesis of the legitimacy of the behavioral representation of latent traits and has already had other designations, such as intrinsic, factorial, and face validity [68]. Construct validity can be analyzed

**Fig. 2** Taxonomy definition method



from several angles, from Classical Measurement Theory (TCM) and Item Response Theory (IRT) [71].

In our model, the latent trait is the support for awareness and collaboration provided by the collaborative environment. [100] list as procedures that identify the latent trait, factor analysis, correlation with other tests, internal consistency, and convergent and discriminant validation.

The validation of the proposed model was carried out in two stages.

First, to improve the proposed assessment model, we expose the model’s artifacts to expert appreciation through

the expert panel approach [12] (see Sect. 5). In this scenario, we seek to expose our taxonomy and assessment model artifacts to the scrutiny of experts to collect an accurate model’s criterion and content validity. The expert panel is composed of a multidisciplinary group of senior researchers with backgrounds in computing or statistics. The review analyzes the usefulness aspects, namely, clarity, relevance, consistency, and completeness of the measurement instrument items.

After this refinement, we reviewed the exposed artifacts; then, we started the large-scale model evaluation process by planning and executing a case study [104, 106] through a



large-scale evaluation of the assessment model (see Sect. 6). Based on the results obtained, we evaluate the proposed model regarding reliability (Sect. 7) and dimensionality (Sect. 8).

Data regarding reliability and construct validity were analyzed following the definition of [99] and the scale development guide [25]. For reliability measurement, we considered the internal consistency through Cronbach's alpha coefficient [20]. Exploratory Factor Analysis (EFA) and Confirmatory factor analysis (CFA) were applied to test dimensionality [45, 50, 71].

## 4 The awareness assessment model

The assessment model comprises the Awareness Assessment Process and the Conceptual View. A full reference is available at the assessment model repository [60].

This model was developed specifically for evaluating collaborative systems by measuring the quality of awareness information supported. At least one examiner conducts the assessment. Considering the participants' perception as a data source, this instrument allows us to classify the collaborative environment into the awareness quality level.

### 4.1 The awareness assessment process

Our awareness assessment process is based on a set of HCI guidelines [10, 80] and is inspired by the evaluation process defined by the standard ISO/IEC 25040:2011 [32]. The assessment process comprises three phases: planning, execution, and reflection (see Fig. 3).

*Phase 1—Planning.* It refers to activities related to assessment planning and involves three basic steps: determine the assessment objectives, the assessment objectives, and the planning assessment.

First, the examiner determines the assessment objectives. This is the starting point for building the evaluation approach and aims to select three essential activities: assessment objectives, context, and goals.

*Activity 1.1. Define the assessment objectives.* This step defines the evaluation goal in terms of the object of study, purpose, perspective, and context [11]: the purpose defines the intention of the evaluation; the perspective tells the viewpoint from which the evaluation results are interpreted (e.g., users or experts); and the context is the environment in which the evaluation is performed.

*Activity 1.2. Select the awareness dimensions.* Identify the related awareness dimensions that will be considered in the assessment. The complete assessment model consists of three primary awareness perspectives and allows us to assess the collaborative environment from each perspective.

*Activity 1.3. Select the goals to be measured.* For each awareness dimension considered, select which design categories are relevant in the collaborative environment. These design categories represent the specific awareness assessment goals, thus allowing the flexibility of the model to address the relevant aspects of the application.

Second, the examiner determines the scope. This phase represents the detailing of the context in which the evaluation will be carried out, the features of the environment that will be considered, the participants, their respective tasks, and finally, whether the boundary, persona, and historical implications will be considered.

*Activity 2.1. Select the features to be evaluated.* Select the features or tasks of the collaborative environment to access. In some cases, the target environment can be complex, thus making it difficult to assess it thoroughly, and some parts/features are not interesting for the intervention. This activity allows building an assessment instrument focused on the relevant/exciting aspects.

*Activity 2.2. Define participants and tasks.* Identify the participants involved in the evaluation process and the tasks that must be carried out within the collaborative environment. This evaluation instrument was designed to enable evaluation by specialists involving users or even both. Thus, clarifying who is involved and their tasks in the environment is vital.

*Activity 2.3. Identifies the Boundary, Persona, and Historical implications.* The awareness information can be categorized in the perspectives of Boundary, Persona, and Historical Awareness.

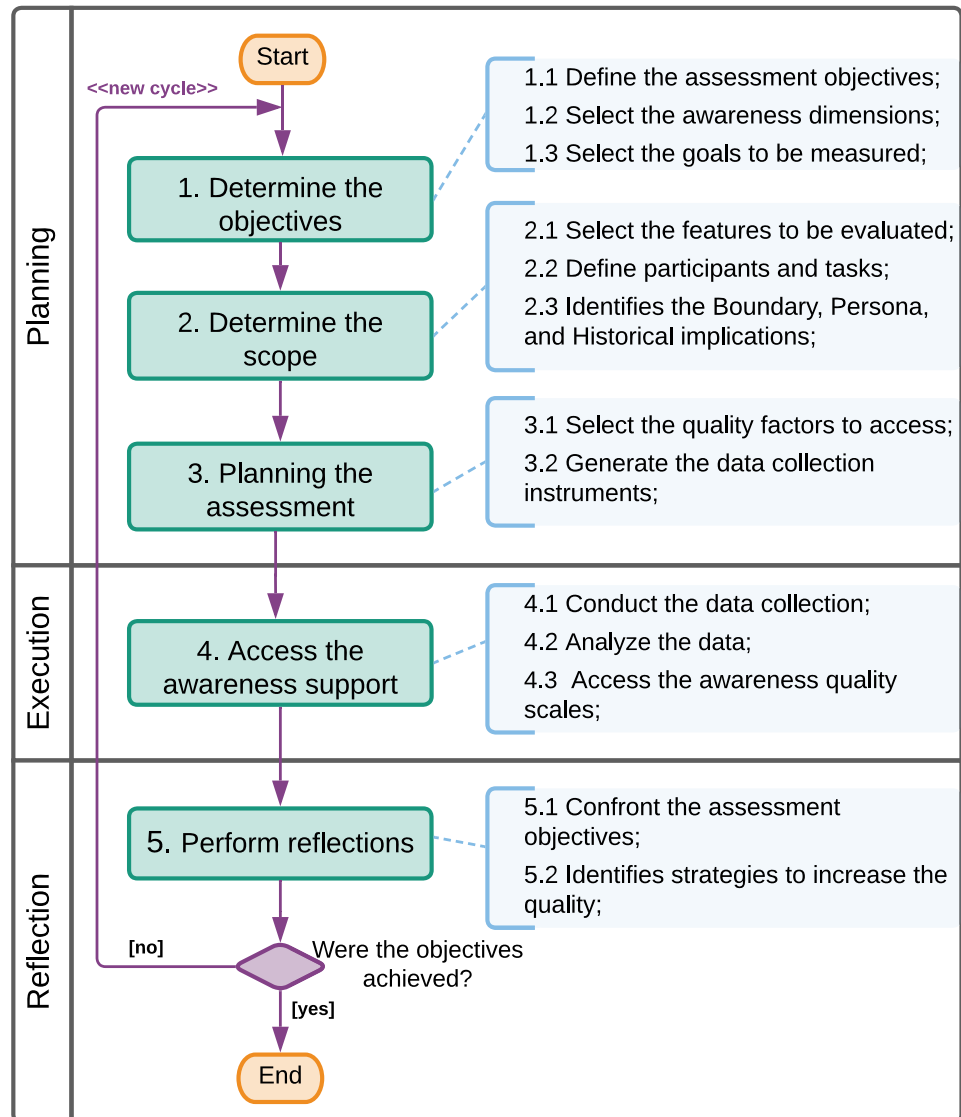
Third, the examiner determines the planning assessment. This phase represents the planning stage documentation, where it is established at what moment of the construction or use of the collaborative environment the evaluation will be carried out and which quality factors will be considered. Therefore, the data collection instrument is prepared (see Sect. 4.4), including the assessment purpose, methods, life cycle, and artifacts.

*Activity 3.1. Select the quality factors to access.* The quality aspects define the additional quality factors under analysis in the evaluation, that is, to use this model together with another evaluation approach (e.g., usability, demographic, user experience).

*Activity 3.2. Generate the data collection instrument.* This activity aims to prepare or customize the data collection instrument, considering the raised in activities 1.2, 1.3, and 2.3.

*Phase 2—Execution.* After the planning stage, the collaborative system assessment is done by adopting data collection and analysis instruments. In this phase, the examiner performs the awareness assessment of the target collaborative environment. The awareness support

**Fig. 3** Awareness assessment process



provided by the target environment is reached through the data collection and analysis tools.

*Phase 3—Reflection.* Once the evaluation is completed and the data is analyzed, the evaluator conducts reflections to gather feedback and identify strategies for improving awareness quality. The main objectives are confronted, and the awareness of quality factors is checked. If unmet, the examiner determines strategies to increase the awareness mechanisms quality indicators, and a new intervention can be planned. This process enables both the assessment of collaborative environments through awareness mechanisms and the improvement by prompting reflection on results.

## 4.2 The conceptual view

The Conceptual View is a framework composed of the following artifacts:

- (1) The awareness taxonomy is constituted of three main awareness dimensions, their respective design categories, and respective design elements, combined with three additional dimensions that directly imply the design categories and awareness elements: persona, boundary, and historical awareness dimensions (Sect. 4.3); full reference can be found at [61];
- (2) The assessment planning protocol represents an instrument for planning and executing the assessment process. This artifact helps in defining the assessment objectives, factors to be measured, awareness dimensions, life-cycle phases in which the awareness assessment will be applied, and so on (Sect. 4.4);
- (3) The data collection and analysis tools present a set of support artifacts for conducting the collection and compilation of data obtained by interventions (Sect. 4.5);

(4) The assessment scales and measurement items represent useful elements to analyze and classify the collaborative environment at an awareness quality level through the participants' perspective (Sects. 4.6).

### 4.3 The awareness taxonomy

As a starting point, we consider the contributions of Gutwing's 5W+1 H awareness framework [43] combined with the awareness classifications of Antunes et al. [5] and Do Espírito Santo et al. [26].

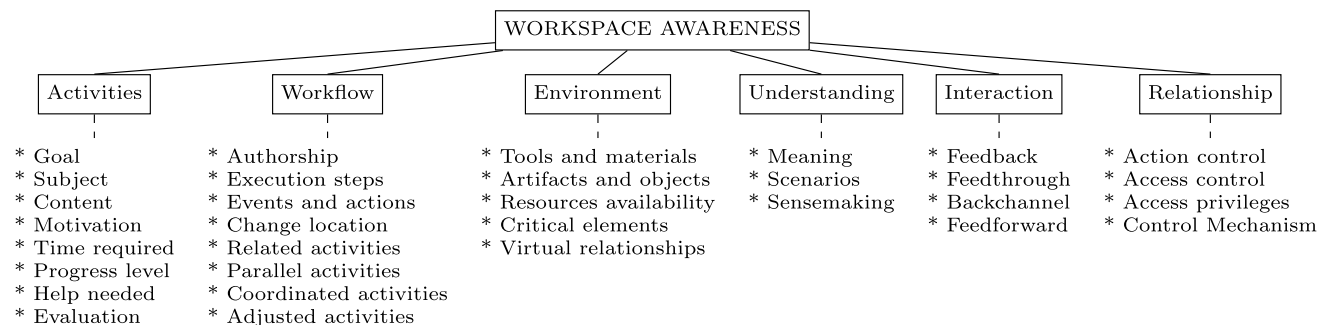
Our taxonomy presents three main awareness dimensions in collaborative applications: workspace, collaboration, and contextual awareness (see Fig. 4). The awareness dimensions consist of a 4-level hierarchical representation structure, containing the awareness dimension, their design

categories, design elements/awareness mechanism involved, and the 5W+1 H framework correspondence [43].

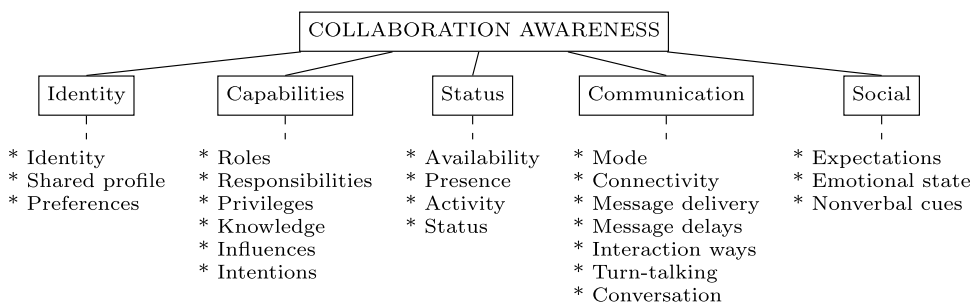
#### 4.3.1 Workspace awareness

The workspace awareness category is a virtual container of places where members can share artifacts, objects, tools, and materials with others. It represents a set of ongoing activities and allows members to interact with each other. This view contains six design categories and 32 design elements, as presented in Fig. 4a.

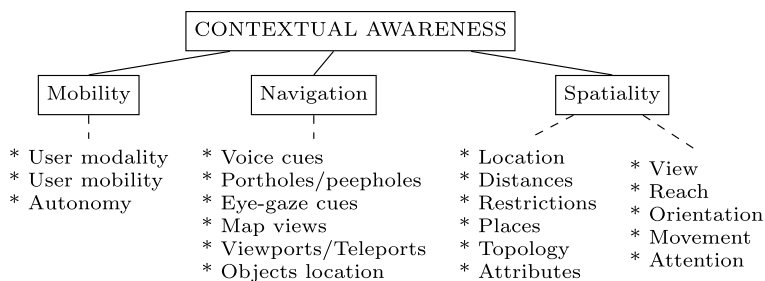
*Activities* design category refers to activities, tasks, objects, and other elements existing in the shared environment. It contains eight design elements: *Goal*, the larger activity or goal that an action contributes to; *Subject* or artifact that is being altered; *Content* up to date; *Motivation* for



(a) Workspace awareness category



(b) Collaboration awareness category



(c) Contextual awareness category

Fig. 4 Taxonomy overview

actions taken; *Time required* to perform the tasks; *Progress level* in carrying out activities and goals or tasks done by the group; *Help needed* to complete the tasks; and *Evaluation* of results.

*Workflow* design category refers to a global perception of steps involved in a working process. It contains eight design elements: *Authorship* of actions being carried out in the environment; *Execution steps* the activities or steps necessary to complete the objectives that provide indications on how a task is being (or was) carried out, and shows activities being performed by a particular user; *Events and actions* that occurred in collaborative environment as a way to help users understand what is happening, providing information on group progress on accomplishment of project tasks, actions performed by participants individually, and understanding of actions performed by others as a group over time; *Change location* that indicates the place where a user is currently working on; *Related activities* that give us information about project-related activities of group members; *Parallel activities* being performed by users; *Coordinated activities* being performed by users (e.g. through a workflow); and *Mutually adjusted activities* being performed by users (e.g., modifying their own work according to others' activities).

*Environment* design category contains information regarding the space used or required for work and its resources. It contains five design elements: *Tools and materials* required to tasks; *Artifacts and objects* in the workspace, such as information on changes performed on artifacts created by the group or information about group members' actions on artifacts created by the group; *Resources availability* that indicates whether a resource is shared for a group, public, or private; *Critical elements* that highlights the presence of critical issues in the working environment (e.g. events or situations); and *Virtual relationships* between objects/resources in workspace.

*Understanding* design category provides insights into what is happening and how individual, coordinated, and collaborative efforts influence group decision-making. It contains three design elements: *Meaning* about what is happening in the working environment; *Scenarios* or cues about future situations that may occur in the working environment; and *Sense making*, being individual, distributed, collaborative, and general. Individual sense-making represents information that helps users reflect on their course of action. Distributed sense-making is cues regarding environmental changes that may be relevant to the action. Collaborative sense-making constitutes information that helps users keep a shared sense of their goals and achievements. General sense-making represents an understanding of other participants and their objects.

*Interaction* design category represents responses from individuals, others, or group actions through a groupware system that allows users to understand the effects resulting

from the interaction. It contains four design elements: *Feedback* about user's current actions; *Feedthrough* about other people's current actions; *Backchannel feedback* notifies the user if others are following what she/he is doing; *Feedforward* indicates updates of in-progress tasks.

*Relationship* design category represents the relationship and dependency between activities, tasks, or shared objects, rules, precedence, or constraints imposed to their realization. It contains four design elements: *Action control* over each user's actions and decisions; *Access control* about who is in control of a shared object/resource; *Access privileges* of data or group activities; and *Control mechanisms* that indicates whether an access control mechanism is being used (concurrency control, floor control, version control).

#### 4.3.2 Collaboration awareness

The collaboration awareness category refers to the perception of the group's availability, structure, and interaction aspects. It was categorized into five design categories and contains 23 design elements, as presented in Fig. 4b.

*Identity* design category composes the individual profile and contains three design elements: identity: *Identity* of the people that are interacting with a system; *Shared profile* with other people; and *Preferences* of group members and the group as a whole.

*Capabilities* design category involves the participants' skills, knowledge, and assumptions that help them to outline their respective roles and to design the cooperative work. It contains six design elements: *Roles* that people and system can have; *Responsibilities* of participants; *Privileges* what participants can do; *knowledge* of the state of an environment; *Influence* level that people can have; and *Intentions*, plans or motivations of those people.

*Status* design category presents information that allows monitoring the current situation/availability of the participants, system, task, and environment. It contains four design elements: *Availability* of group members; *Presence* of people over time; *Activity level* of the user engaged in his device; and *Status* the current system setup and the state of the interface.

*Communication* design category is related to information that guides participants in establishing and managing communication channels for interacting with others. It contains seven design elements: *Mode* (Synchronous/Asynchronous), indicating whether other users are working online, offline, or both; *Network connectivity*, indicating whether the user is connected or not; *Message delivery*, the target users receive message notifications; *Message delays*, information on the time spent in message delivery; *Interactions ways* that allows peers to establish links with each other; *Turn-talking*, who is talking, who is listening, whose ideas it is and whose turn to speak; and *Conversation* with other participants.

*Social* design category represents information on the participants' social perspective, emotions, expectations, premises, or other nonverbal cues that help to understand their actions, attitudes, or behavior. It contains three design elements: *Expectations* about other group members; *Emotional* state of the participants; and *Non-verbal cues* about social information.

### 4.3.3 Contextual awareness

The contextual awareness category represents the notion of physical and virtual spaces, their topology, interaction ways, and mobility issues. It allows the group to maintain a sense of what is happening in virtual space and covers concepts of group navigation, physical/virtual spaces, spatiality, and mobility. This view comprises three design categories and contains 20 design elements, as presented in Fig. 4c.

*Mobility* design category consists of elements that help users to move from one position or situation to another, usually a better one, whether this situation is related to the device, user, or even real/virtual environment. It contains three design elements: *User mobility*; *User modality*; and *Autonomy*.

*Navigation* design category represents information that assists participants through the shared environment. It contains six design elements: *Voice cues* that provides feedback about who is talking to whom; *Portholes/peepholes* to preview some contents without having to access them; *Eye-gaze cues* about where users are looking at; *Map views* or other visual information from a remote environment; *Viewports/teleports* for users to peek others' activities; and *Objects/Artifacts location* that allows users to identify and share objects/resources.

*Spatiality* design category represents information on the user's physical and virtual perspective and assists users in locating themselves in a shared environment. It contains 11 design elements: *Location* of each participant over time (whether the user is in the same or another place); *Distances* of the user to others; *Constraints* imposed by the physical environment (e.g., object/resource constraints such as location or ownership); *Places*, both physical (e.g., meeting rooms and cafeteria) and virtual (e.g., different places for collaboration); *Topology* of virtual environment (e.g., moving between virtual places) that give cues about the complexity of physical environment where it is used; *Attributes* of objects/resources in the workspace or environmental conditions of place where it is used (e.g., weather conditions); *View* where participants can see; *Reach* represents where participants can reach; *Orientation* of other users; *Movement*, direction, and speed of a user regarding other users; and *Range of attention* when performing activities.

### 4.3.4 Additional perspectives

We established three additional dimensions directly implying the design categories and awareness elements: persona, boundary, and historical awareness. This representation allows us to typify, for each aspect of the awareness taxonomy, the awareness information/mechanism itself, which role this information belongs to, at what time this information represents, and, for contextual ones, their spatial origin.

The Boundary dimension indicates if awareness elements belong to the physical or virtual context (*where?*). The Persona dimension indicates to whom the awareness information belongs (*who?*), allowing the classification of the awareness elements among individuals, other participants, the group as a whole, or groupware/system perspective. The historical awareness dimension represents the temporal information (*when?*)—past, present, and future) carried out during collaborative work, whether situational, contextual, or workspace information.

In the literature, historical awareness is considered awareness information; however, in our understanding, the historical perspective is broader and encompasses all other existing awareness elements in the taxonomy. A detailed overview of these additional perspectives can be found in [61] and in the supplementary taxonomy materials available in the model repository [60].

## 4.4 The assessment planning protocol

The planning protocol template (Table 4) consists of a form that describes the procedures performed in the planning stage: determining the intervention's objectives, scope, and life cycle. An example of a planning protocol is available in Table A7; a usage scenario is detailed in Sect. 6.

## 4.5 The data collection tools

The awareness assessment model contains a set of data collection instruments by applying ten awareness assessment questionnaires. We considered adopting 75 specific awareness assessment items identified in the awareness taxonomy [61].

To reduce the number of assessment items for each participant, we recommend using the balanced incomplete block design approach [48]. A Balanced Incomplete Block (BIB) consists of treatments  $t$  (a subset of the assessment items) that appear in the same block  $b$  (questionnaire) with each other treatments the same number of times  $\lambda$ . The BIB design must satisfy the following characteristics [48]:

- (1) Each block  $b$  have the same number of plots  $k$  (treatments), where  $b.k = t.r$  and  $b \geq t$ ;

**Table 4** Planning protocol template

1. Determine the objectives	Step 1.1. Define the assessment objectives regarding the object of study, purpose, perspective, and context → <i>Object of study:</i> → <i>Purpose:</i> → <i>Perspective:</i> → <i>Context/environment:</i> Step 1.2. Select the awareness dimensions. Identify the related awareness dimensions that will be considered in the assessment → <i>Awareness dimensions:</i> Step 1.3. Select the goals to be measured. We select which design categories are relevant in the collaborative environment for each awareness dimension → <i>Goals:</i>
2. Determine the scope	Step 2.1. Select the features to evaluate. Select the functionalities or tasks within the collaborative environment that will be the object of the assessment → <i>Features:</i> Step 2.2. Define the participants involved and the tasks that must be carried out within the collaborative environment → <i>Participants:</i> → <i>Tasks description:</i> Step 2.3. Identifies the Boundary (physical, virtual, or both), Persona (individual, other participants, group as a whole, or system), and Historical (past, present, future) implications → <i>Implications:</i>
3. Planning the intervention	Step 3.1. Select the additional factors to access (like demographic, usability, and UX) → <i>Additional factors:</i> Step 3.2. Generate the data collection instrument considering the activities 1.2, 1.3, and 2.3

- (2) Every treatment is replicated  $r$  times in the design, where  $r > k$ ;
- (3) Each treatment occurs at most once in a block, and every pair of treatments occurs together  $\lambda$  times in the blocks, where  $\lambda(t - 1) = r(k - 1)$ ;
- (4) Variables  $b, t, k, r$  and  $\lambda \in \mathbb{Z}^+$ .

To satisfy these relationships, we adopted the values of  $b = 10$ ,  $t = 5$ ,  $k = 2$ ,  $r = 4$ , and  $\lambda = 1$ . In this setup, the 75 awareness assessment items were grouped into five blocks of 15 assessment items each. Hence, we used questionnaires containing two blocks of items, totaling 30 questions.

Applying the BIB method, we found a balanced incomplete block design composed of 10 blocks (questionnaires). Table A1 in Appendix A presents the configuration of the treatments ( $t$ ) and blocks of questionnaire items ( $b$ ).

The assessment model comprises ten specific questionnaires representing the main awareness dimensions existing in collaborative environments. It was developed based on a multidimensional perspective represented by three main awareness dimensions (collaboration, workspace, and contextual). The questionnaires were composed similarly for each awareness dimension.

The assessment instruments were developed using the guidelines presented by [11, 25, 103, 104]. In each applicable evaluation question, the participant selects an option according to how much the participant agrees or disagrees with each statement (gradual scale).

Tables A2 to A5 in Appendix A present the complete version of the questionnaire. The questions were composed by combining the following structure: the component of the sentence (subject + predicate), presented in Table A2, combined with the correspondent complement of each awareness assessment item (Table A3, A4, and A5).

#### 4.6 The awareness assessment scale

The awareness quality scales and awareness mechanisms measurement aims to classify the collaborative environment at a quality level through the participants' perspective. The awareness mechanisms measurement allows us to assess the general awareness quality of the collaborative environment, its presented design elements, goals, and awareness dimensions by estimating the examinee's ability.

In this sense, we assume the graded item response approach combined with the ability and item information functions proposed by [8, 81]. The quality scales have been developed adopting the Item Response Theory (IRT) model [7]. The IRT refers to a family of mathematical models that relate observable variables (questionnaire items) and hypothetical unobservable traits or aptitudes (awareness quality).

The IRT model establishes a link between the properties of items on an instrument, individuals responding to these items, and the underlying trait being measured. Thus, a stimulus (item) is presented to the subject, and he/she responds to it, and the response that the subject gives to

the item depends on the subject’s level in the latent trait or ability [70].

We assume that the latent construct (e.g., stress, knowledge, attitudes) and items of a measure are organized in an unobservable continuum. At each ability level ( $\theta$ ), there will be a probability  $P$  that an examinee with that ability will correctly answer the item [9]. The function of ability  $P(\theta)$ , also represented by the item characteristic curve (ICC), describes the relationship between the probability of a correct response to an item and the ability scale.

To calculate the  $P(\theta)$ , we assume the gradual response model presented by [81], where we believe that an item’s response categories can be ordered with each other. On the model, the probability of a participant  $j, \forall j \in J = \{1, 2, \dots, m\}$  chose a score  $k, \forall k \in K = \{0, 1, \dots, n\}$ , for a measurement item  $i, \forall i \in I = \{1, 2, \dots, o\}$  is given by Eq. 1.

$$P_{i,k}(\theta_j) = \frac{1}{1 + e^{-a_j(\theta_j - b_{i,k})}} - \frac{1}{1 + e^{-a_j(\theta_j - b_{i,k+1})}} \tag{1}$$

where  $P_{i,k}$  is the probability that an individual  $j$  receives a score  $k$  in item  $i$ ;  $e$  is the Euler’s number (equals to 2.71828 ...);  $m, n$  and  $o$  are respectively, the total of participants, item scores and measurement items; and  $b_{i,k}$  is the difficulty parameter of the  $k$ -th category of item  $i$ , considering  $b_{i,1} \leq b_{i,2} \leq \dots \leq b_{i,n}$ .

Each item in a test will have its item characteristic curve, and we considered two technical properties to describe it: the discrimination ( $a$ ) and the difficulty ( $b$ ). The discrimination parameter defines how well an item can discriminate (differentiate) the participants about the latent trait (awareness quality), where the higher its value is, the more associated with the latent trait is the questionnaire item. The difficulty parameter indicates the category of the scale in which the item has more information, i.e., where the item functions along the ability scale.

The items’ discrimination is to be interpreted by [7]. A measurement instrument item is satisfactory in a measurement scale if the discrimination value  $a \geq 0.65$ , as presented in Table 5. Thus, measurement instrument items with a discrimination parameter  $a < 0.65$  are disregarded from the analysis, as they may not correctly differentiate the quality level.

Based on the parameters of discrimination and difficulty, it is possible to interpret how the measurement instrument items contribute to the definition of a measurement scale. To position the items on the scale and identify the categories of the scale (quality levels), the model considers the probability parameter  $P_{i,k}(\theta) \leq 0.5$  and scale (0, 1) [22].

IRT widely uses this scale to represent, respectively, the mean value and the standard deviation of the individual abilities of the population. In this case, the values of parameter  $b$

**Table 5** Item discrimination

Classification	Range of values
Very low	< 0.34
Low	0.35 to 0.64
Moderate	0.65 to 1.34
High	1.35 to 1.69
Very high	> 1.7

vary between  $-2$  and  $+2$ . Regarding the parameter  $a$ , values between 0 and  $+2$  are expected; the most appropriate values are those greater than 0.65.

### 4.7 The awareness measurement mechanisms

On the IRT, the evaluation information is defined in terms of item information functions  $I_i(\theta)$ , which is a measure of how sound responses in that category estimate the examinee’s ability [8].

Our model assumes the graded item response approach, where each item has been divided into  $n$  ordered response categories. Then, for each awareness dimension  $d, \forall d \in D = \{workspace, collaboration, contextual\}$  and considering the applicable awareness goals  $g, \forall g \in G_d = \{\forall g \mid g \text{ is a goal} \in \text{awareness dimension } d\}$ , their related measurement items  $i, \forall i \in I_{gd} = \{\forall i \mid i \text{ is a measurement item} \in \text{awareness dimension } d \text{ and goal } g\}$ , and item scores  $k$ , denoting an arbitrary category  $\forall k \in K = \{0, 1, \dots, n\}$ , where  $n$  is the number of response categories for item  $i$ , the model calculates:

- (1) The item’s information  $I_i(\theta)$ , for each applicable questionnaire item  $i$ , considering the awareness goal  $g$  of the awareness dimension  $d$ , using the item information function proposed by [81] (Eq. 2).

$$I_i(\theta) = \sum_{k=0}^n \frac{[P_{i,k-1}^*(\theta) - P_{ik}^*(\theta)]^2}{P_{i,k-1}^*(\theta) - P_{ik}^*(\theta)} \tag{2}$$

where,

$$\sum_{k=0}^n P_{ik}(\theta) = 1$$

- (2) The awareness goal’s information  $GI(\theta)$ , for each applicable goal  $g$  of the awareness dimension  $d$ , that is calculated considering the information for all applicable items  $I_j(\theta)$  using the test information function presented by [9] (Eq. 3), where  $m$  is the number of applicable items of goal  $g$ ;  $I_j(\theta)$  is the item’s information for each applicable goal item.

$$GI(\theta) = \sum_{j=1}^m I_j(\theta) \quad (3)$$

- (3) The awareness dimension's information  $AI(\theta)$ , for each applicable awareness dimension  $d$ , is calculated considering all applicable goal  $g$  using the test information function presented by [9] (Eq. 4), where  $o$  is the number of related goals  $g$ ;  $GI_j(\theta)$  is the amount of information for each applicable goal.

$$AI(\theta) = \sum_{l=1}^o GI_l(\theta) \quad (4)$$

The Eqs. 2, 3, and 4 calculate the information scores from a single participant viewpoint; thus, to transfer these values to the collaborative environment perspective it is necessary to calculate the average of the provided scores  $I_i(\theta)$ ,  $GI(\theta)$ , and  $AI(\theta)$ , considering all participants involved.

## 5 Expert panel validation

To improve the assessment model, the model's artifacts were exposed to expert appreciation through the expert panel approach [12]. Based on the Goal Question Metric approach [11, 102], we designed an evaluation questionnaire by decomposing the study objective into quality aspects and analysis questions. The expert evaluation questionnaire contains three demographic questions and ten assessment items related to the usefulness concept, as presented in Appendix A Table A6. The supplementary materials are available at [60].

This review aims to analyze the usefulness aspects, namely, clarity, relevance, consistency, and completeness of the measurement instrument items from the researchers' perspective. The usefulness is related to the purposeful, unambiguous determination and applicability aspects [66].

Purposeful is the significance and objectivity of the model and its elements. Unambiguous determination is the ability to represent its elements and characteristics clearly, concisely, and unambiguously. Applicability seeks to assess its practical use for classifying, differentiating, and comparing objects.

In this context, the expert panel validation allows us to address whether a purposeful and unambiguous determination is possible by evaluating the practical applicability and demonstrating whether a clear definition of its elements can be made [90]. This approach also allows reflecting on the current state of research on an object [53], to discover

similarities and differences between studies on this type of object [1], and to identify potential research gaps [49].

### 5.1 Expert panel results

In this step, we exposed the assessment model to expert evaluation, like awareness, collaborative systems, and HCI researchers, to identify its suitability for evaluating collaborative environments. After this refinement, we reviewed the exposed artifacts and started the large-scale model evaluation process through a case study.

We presented the data collection artifacts (questionnaire) to expert opinion, and five expert assessments of the initial model were obtained. Overall, the evaluation model received a good rating from the expert's perspective. Figure 5 summarizes the obtained results.

On a gradual scale, from 1 (strongly disagree) to 5 (strongly agree), the assessment items M1 to M7 received values over 3,5 (average 3,8). Despite the small sample of specialists, all reported having good experience regarding key concepts of awareness (D1), collaboration (D2), and HCI (D3), corroborating the quality of the responses. On a gradual scale, from 1 (novice) to 5 (expert), the reported expertise was close to 5 (average 4,1).

Regarding understandability and completeness (M3 and M7), the feedback received demonstrates a concern regarding the clarity of the specification and whether the model contains all statements about the domain or can be applied to the same environment.

We thought that, depending on the domain of the collaborative system, not all aspects would be applied—and this will not necessarily be a weak point of the model. For example, the awareness information may differ if a system focuses on performing synchronous or asynchronous work. In some points, the awareness mechanisms require balancing

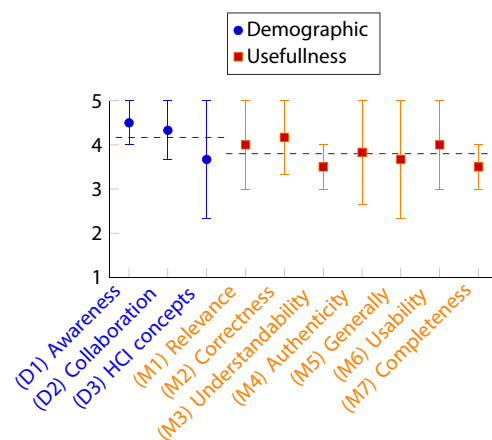


Fig. 5 Expert panel questionnaire results



the need to present proper awareness support while dealing with information overload or intrusiveness.

We considered the feedback obtained through the expert panel to refine the assessment instrument, especially regarding the item syntax. We also adjusted the items' scale to 4 points, as we believe that a neutral position, as occurs in a 5-point scale, does not corroborate with our intended analysis model. After the expert panel refinement, we reviewed the exposed artifacts and started the model evaluation process by planning and executing a case study.

## 6 Case study validation

In this scenario, virtual collaboration environments intended for simultaneous communication or interaction between two or more people were evaluated, for example, conference environments, videoconferencing, virtual events, webinars, etc. In these environments, for a satisfactory exchange, it is necessary to provide awareness cues such as the participants' profile, capabilities, status, communication ways, and social aspects.

Initially, we planned the case study using the awareness assessment process (described in Sect. 4.1). As a result of this step, the planning protocol artifact was created, where the intervention's objectives, scope, and life cycle were determined.

We compiled the blocks (treatments) into ten different test books and then set up a printed questionnaire and an online version (Google Forms) to collect participant feedback. The questionnaires were prepared as described in Appendix A, Tables A2 to A5. The full version of the case study materials and the IRT dataset is available at [60].

### 6.1 Model calibration

After applying the questionnaires, all observations were compiled into a.csv file. To calibrate the assessment model, we ran the IRT script available at the assessment model repository [60] and interpreted the output values of discrimination ( $a$ ) and difficulty ( $b$ ) disregarding items with  $a < 0.65$  or  $a > 4.0$  (as defined in Table 5).

We analyzed the observed frequencies of each response category for all questionnaire items and grouped those with a small number of responses [45] ( $< 10$  observations for each category). In these cases, the response categories "strongly disagree" with "disagree", or even the categories "agree" with "strongly agree" were combined. Then, we re-run the model with the remaining items and generated the final discrimination and difficulty coefficient.

The workspace awareness assessment items (see Table A3), Q1—goal, Q2—subject, Q3—content, and

Q30—access control were removed from the calibrated model version, as they did not present values compatible with the range defined for the parameters  $a$  and  $b$ . In items Q1 to Q3, the observed frequencies indicate that almost all participants could identify this information and mostly assign the category "agree" or "strongly agree" to these assessment items. In item Q30, the values conflicted, and the model did not converge to satisfactory parameters. We conjecture that it may indicate an assessment item strongly linked to user-specific factors or even supported differently in each environment.

From the collaboration awareness perspective (see Table A4), we disregard the assessment items Q33—identity, Q44—activity level, Q46—connectivity mode, and Q48—message delivery in the calibrated model. Participants generally indicated ease in identifying these assessment items and chose answers agreeing with the statement.

From the contextual awareness perspective (see Table A5), the Mobility design category did not present any assessment items converging with the model. Thus, Q67—user modality, Q68—user mobility, and Q69—autonomy were removed from the results. In addition, most of the elements in the category Navigation followed the same criteria. The assessment items Q71 (portholes/peepholes) and Q73 to Q75 (namely, map views, viewports/teleports, and artifacts location) were not relevant to the target scenario, indicating that these resources were absent or had not been used by the participants to collaborate.

Tables A8 to A10 in Appendix A present the coefficients of discrimination ( $a$ ) and difficulty ( $b$ ), the observed frequencies and Cronbach's alpha coefficient ( $\alpha$ ) for the awareness taxonomy items.

The coefficients  $b_1$ ,  $b_2$ , and  $b_3$  are related to each response category. Thus, for the items on the 4-point gradual scale,  $b_1$  represents the 1st category;  $b_2$  represents the 2nd category;  $b_3$  represents the 3rd category; the complement represents the 4th category. For the items where grouping was applied, we used the 3-point gradual scale; therefore, only the parameters of  $b_1$  and  $b_2$  were generated. The NA values represent the cases where grouping was necessary.

### 6.2 Case study results

We obtained the 422 voluntary participation. Regarding gender, we collected 298 male observations (70%) and 112 female observations (27%); 12 participants did not answer this question (3%). We collected 345 observations from individuals aged 18 to 28 years (82%), 58 from individuals aged 29 to 39 years (13%), 17 from individuals between 40 and 50 years old (4%), and two observations of individuals over 50 years old ( $< 1\%$ ). No one under the age of 18 years participated in this research.

We generated frequency histograms based on the participant's scores and demographic facets (Fig. 6). To verify whether the model presents a distinction in discrimination values (or average) of different groups, the model calculates the normal distribution and the mean score grouped by each demographic perspective (Fig. 7).

We collected demographic data such as age, gender, preferred videoconferencing environment, expertise in using collaborative environments, and individual knowledge of collaboration and awareness concepts. The histogram in each demographic facet is mainly within the individual score thresholds where the model is representative (vertical dotted line).

As shown in Fig. 7, the normal curves generated for each group were significantly close, indicating that the model did not present different behaviors in the observed groups and the cumulative probability distribution. Furthermore, the sigmoid function suggests that the model does not significantly differentiate discrimination parameters ( $a$ —the sigmoid slope) and difficulty ( $b$ —the sigmoid midpoint).

In the demographic facet of participants' age, the normal distribution and the sigmoid function do not present a score distortion in 3 of 4 age groups evaluated. The group of young individuals, 18 to 29 years old, gave a slight left-shift in the sigmoid function, demonstrating that, in general, younger people have more straightforward use of these environments compared to older. This factor may have positively

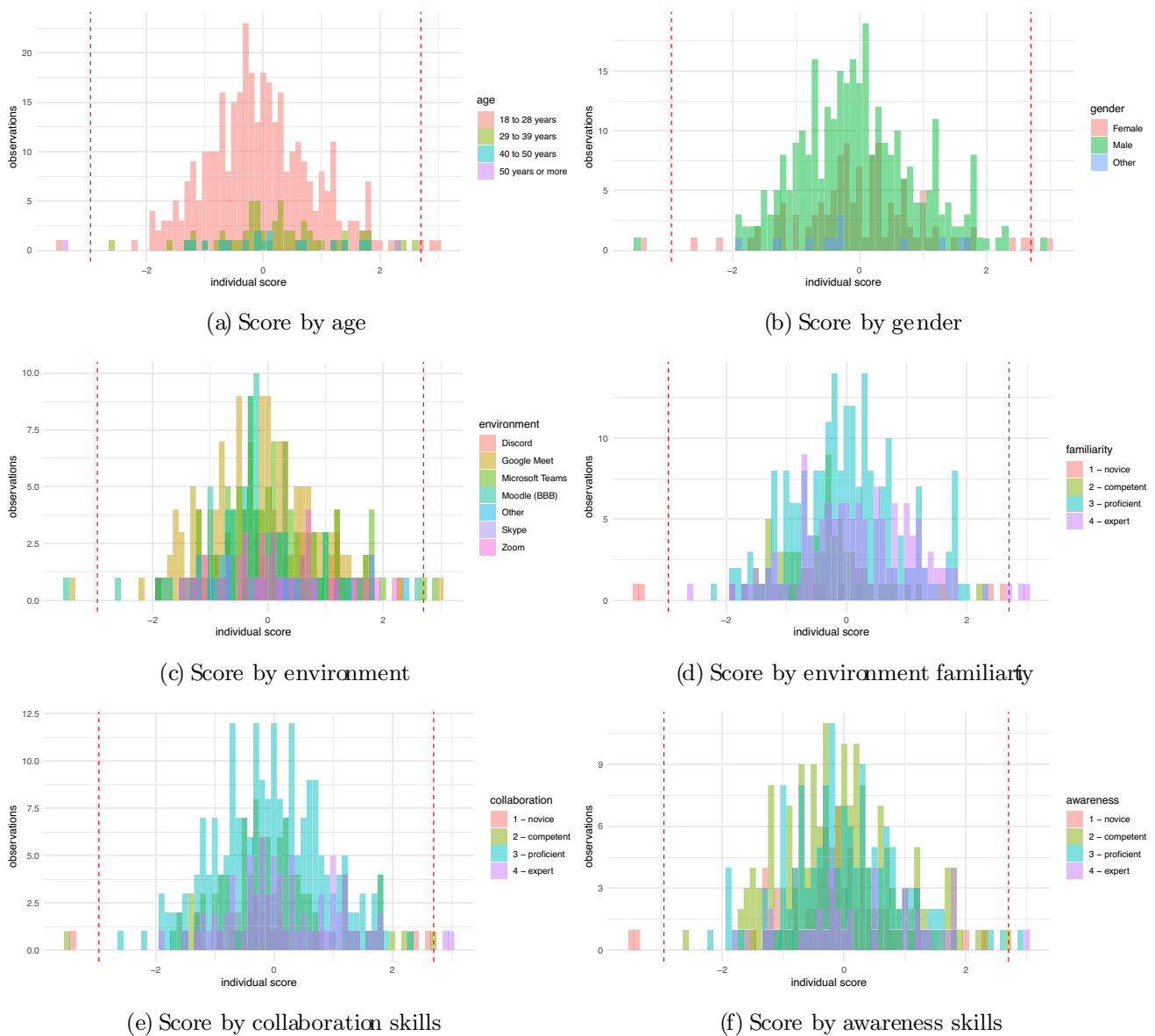
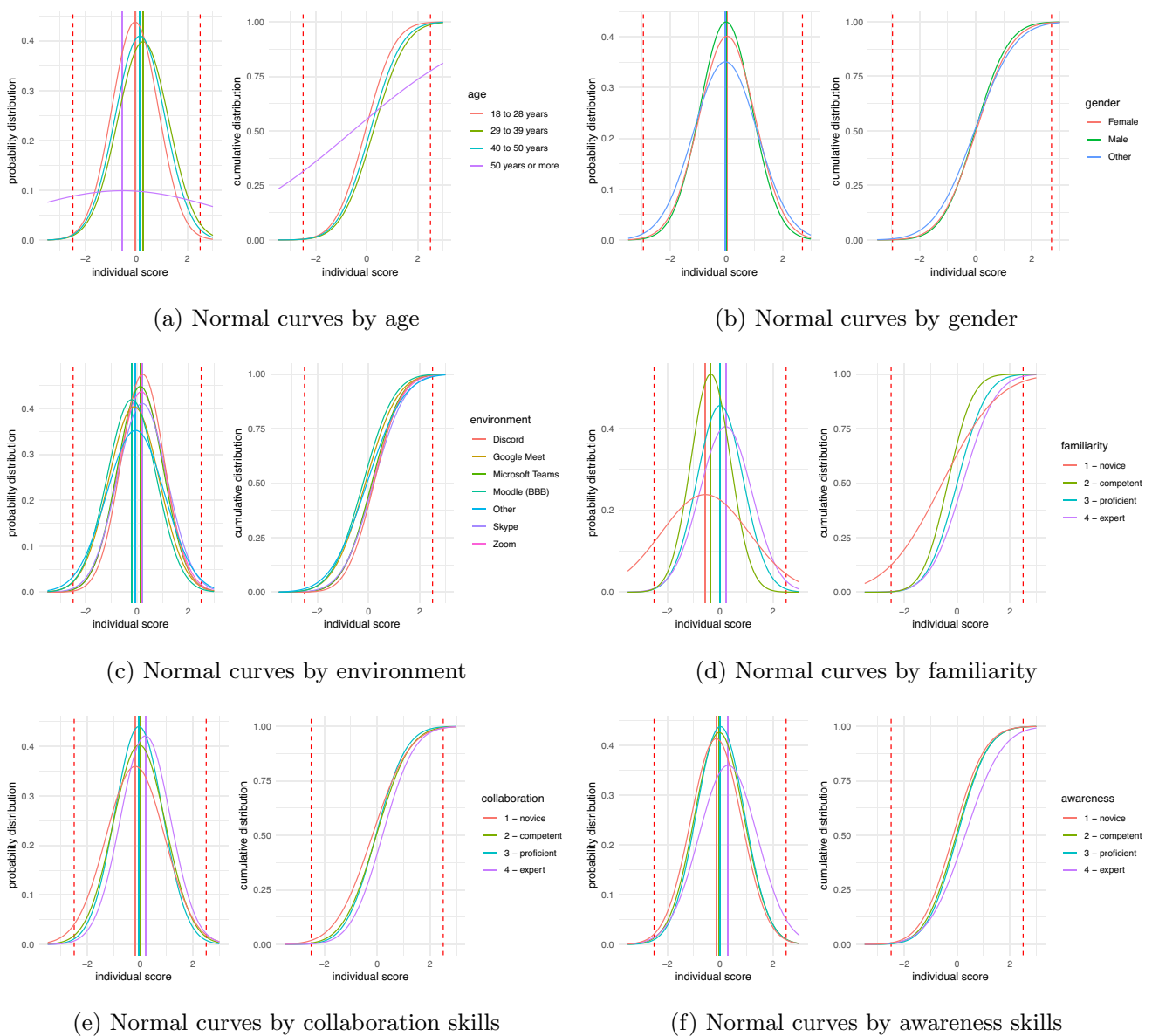


Fig. 6 Demographic distribution of individual score



**Fig. 7** Normal curves of individual score

corroborated the score because the sample of individuals in the first age group was significantly larger. We did not obtain a significant sample of individuals aged over 50 years; thus, the analysis of this group was not possible.

Grouping the participants by gender demonstrated that the model does not present additional difficulties or differentiate participants depending on their gender options. Furthermore, despite the sample mainly being composed of males, females, and other genders, it obtained similar results in both mean scores and normal and sigmoid functions.

Comparing the scores grouped by the preferred videoconferencing environment, we observed that environments Google Meet, Microsoft Teams, and Moodle (BigBlueButton) showed a slight distinction in the average difficulty

parameters (sigmoid curves slightly shifted to the left). This demonstrates that, in general, it was easier to identify the available awareness elements in these environments, and participants performed slightly better than in other environments.

By analyzing the participants' individual skill histograms (Fig. 6d–f), we investigate whether familiarity with the preferred videoconferencing environment, collaboration, or awareness concepts, both normal distribution and probability cumulative distribution (Fig. 7d–f) were compatible with the participant's judgment. The observed frequencies in the histograms indicate a normal distribution for all groups and encompass the entire spectrum of the ability scale.

As shown in Figs. 6 and 7, the model does not differentiate the scale by a specific group of individuals; the factor that distinguishes individuals is, precisely, the latent trait evaluated. In other words, a better participant skill implies better performance on the model scale, which corroborates constructing an appropriate assessment model.

For each awareness mechanism of the taxonomy (described in Sect. 4.3), we also calculated the relationship between the probability of each response item (from strongly disagree to strongly agree) concerning the individual's ability scale. In this representation, the likelihood of the individual evaluating each item considers the difficulty/skill that the participant has, i.e., elements that are more difficult to understand and require a higher skill scale for their assessment.

Figure 8 shows the total information curve of the awareness mechanisms' support and the standard error (SE). The blue line represents the test information function  $I(\theta)$ , represented by a normal (Gaussian) distribution [96]; the red dotted line represents the standard error  $SE(\theta)$ . The intersection point represents the limits at which the model is more representative.

This graph represents the region of the ability scale  $\theta_j$  where the participant  $j$  can access the provided awareness mechanisms. The curve shape indicates that the instrument covers the entire latent trait, from participants who are unable to understand the mechanisms ( $\theta_j < -1$ ) to those who can identify the mechanisms quickly ( $\theta_j > 1$ ).

The total instrument information and SE curves show the instrument's accuracy. The SE curve is observed to reach its minimum value precisely at the point on the scale where the information curve reaches its maximum. Therefore, the instrument is indicated for participants with a skill level in the scale region where the information curve exceeds the standard error curve, interval  $[-2.96, +2.70]$ .

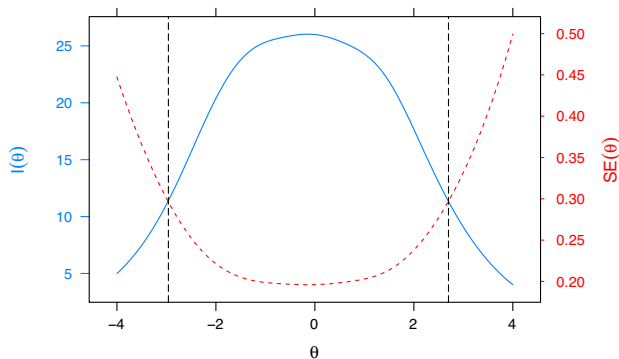


Fig. 8 Test information and SE

### 6.3 The awareness support scale

Applying the awareness measurement formulas (Eqs. 2 to 4) defined in Sect. 4.7, we calculated the probability scales  $P_{i,k}(\theta_j)$  for the assessment element through the IRT awareness assessment model.

The awareness support scale assumes a coverage interval  $[-4.0, +4.0]$ , although this model is representative at the interval  $[-2.96, +2.70]$ , to cover the outlier scores of individuals with lower or higher abilities ( $< 1\%$ ). Figure 9 presents the probability scales generated for each assessment item and awareness dimension.

As exemplified, individuals with lower skill scores generally have more difficulty recognizing the available awareness elements and, therefore, are more likely to disagree with the presence of these elements in the application. On the other hand, individuals with a higher score on the scale are more likely to recognize awareness elements presented by the application and, thus, give more remarkable agreement when judging the items.

For each assessment item, the scale presents the probability of a participant with a given ability score recognizing the available awareness information, and the segments in the graph bars represent the participant's likely response to each statement.

Unlike the score measurement in a standard test of  $n$  right/wrong questions, which generally takes integer values between 0 and 1, in IRT, the participant's ability  $\theta$  can take on any real value between  $-\infty$  and  $+\infty$ . Therefore, it is necessary to establish an origin and a unit of measurement to define the scale [22].

To calibrate the model and construct the graphs shown in Figs. 8 and 9, we considered the scale with a mean  $\mu$  equal to 0, and a standard deviation  $\sigma$  equal to 1. The scale (0, 1) is widely used in IRT to represent, respectively, the mean value and the standard deviation of the individual abilities of the population [22].

Despite the frequent use of this (0, 1) scale, there are no practical differences if these or any other values  $\mu$  and  $\sigma$  are established, as what is important are the order relationships existing between their points. Although this is a standard scale in IRT, its interpretation from the participant's perspective may not be well accepted because an individual with a low ability would have a negative score, which could generate a pejorative connotation.

To overcome a possible scale misinterpretation, we adopted the principle of invariance of the IRT scales [7, 9] and applied a linear transformation (Eq. 5) to establish a more appropriate and easier reference for people to interpret their awareness score through a positive scale  $\theta^*$ .

$$a(\theta - b) = \frac{a}{\sigma}[(\sigma \cdot \theta + \mu) - (\sigma \cdot b + \mu)] = a^*(\theta^* - b^*) \quad (5)$$

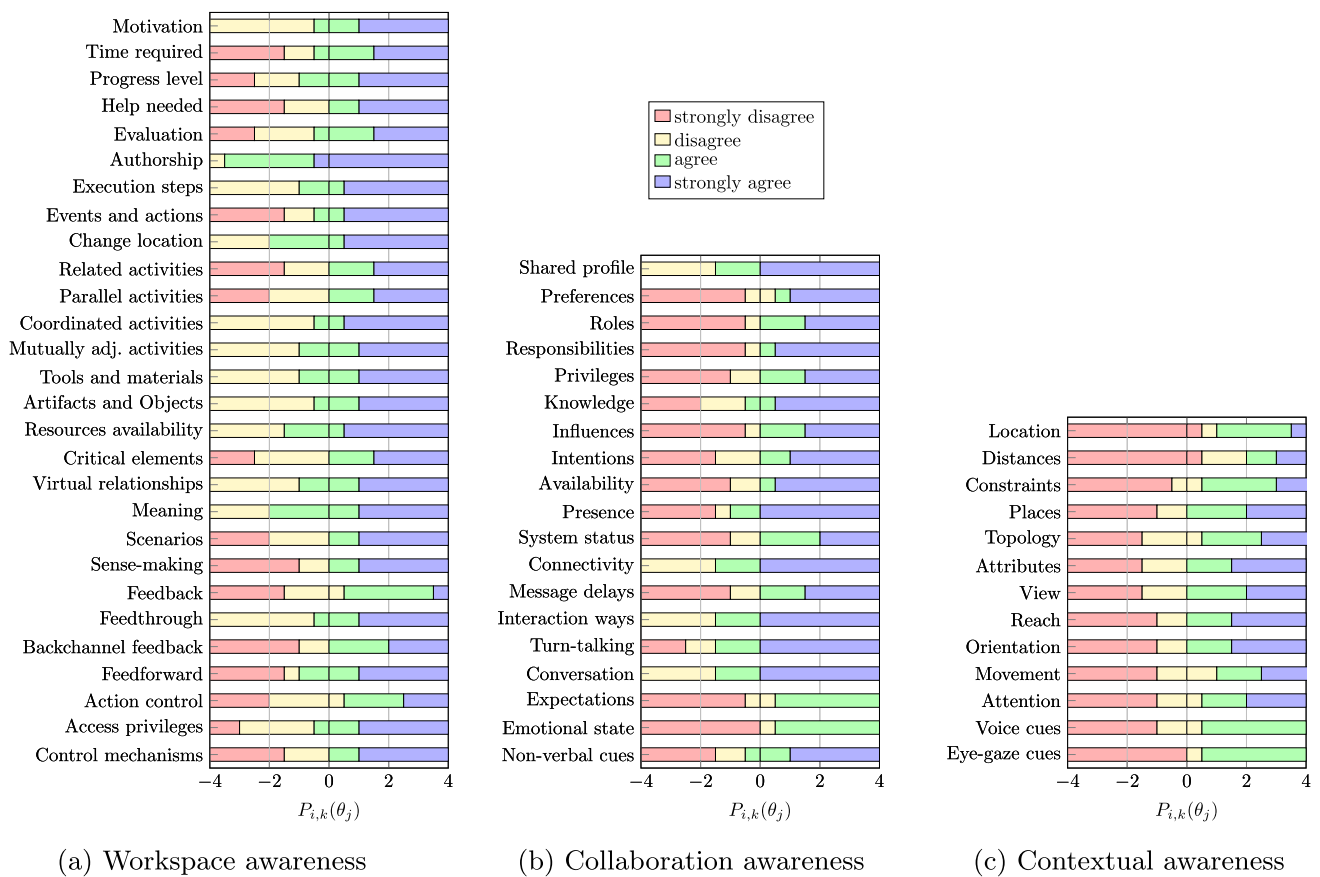


Fig. 9 Ability level scales

with,

$$a^* = \frac{a}{\sigma}; \quad b^* = \sigma.b + \mu;$$

$$\theta^* = \sigma.\theta + \mu; \quad P(\theta^*) = P(\theta)$$

where \* indicates the value at the new scale;  $\mu$  is the mean  $\theta$  value at the first scale;  $\sigma$  is the standard deviation at the first scale;  $\theta^*$  is the adjusted score (the awareness points);

Over the participant’s score  $\theta$  and IRT parameters  $a$  and  $b$ , we applied a linear transformation converting the resultant scores to a new scale (100, 10). In this perspective, the calibrated items were positioned over the awareness scale, establishing three awareness quality levels: low, good, and excellent.

To position the items on the awareness scale and identify the quality levels, we considered the probability parameter  $P_{i,k}(\theta) \geq 0.5$  and the  $\theta$  and  $\theta^*$  scales (awareness points). The awareness quality scale provides an overview of the different profiles of existing users and, by establishing their expected skills, allows us to visualize the likely set of awareness support mechanisms known for each profile and how they perform collaborative activities in the environment.

Tables 6, 7 and 8 present, respectively, the workspace, the collaboration, and the contextual awareness quality scales.

As a resultant process of knowing the awareness profiles and the participants’ scores, like their skills (archived awareness mechanisms) and difficulties (not archived awareness mechanisms), we can trace paths to identify how awareness works and how the collaboration occurs.

Essentially, this model provides reflections toward collaborative improvements by gradually prioritizing supported awareness elements over a participant’s perspective.

### 6.3.1 Scale interpretation

Through the awareness quality scale, we can visualize two complementary facets.

In the first awareness scale perspective, as shown in Fig. 9, we have access to the general performance of the evaluated environments by each assessment item (awareness mechanisms). Thus, for each response category of the IRT gradual scale (from strongly disagree to strongly agree), the model represents the expected ability intervals

**Table 6** Workspace awareness scales

Level	Quality description
Low ( $\theta < -1$ ) ( $\theta^* < 90$ )	The collaborative environment rarely provides workspace capabilities or information about the activities, environment, or workflow. It does not provide the interaction or understanding of artifacts and objects shared in the workspace. Due to these limitations, the interaction is limited
Good ( $-1 \leq \theta \leq 1$ ) ( $90 \leq \theta^* \leq 110$ )	The collaborative environment sometimes supports workspace capabilities and presents information like activities, environment, and workflow. The interaction and understanding of artifacts and objects shared is possible, although it usually does not present good operability
Excellent ( $\theta > 1$ ) ( $\theta^* > 110$ )	The collaborative environment supports workspace capabilities and provides information about the activities, environment, and workflow. The environment provides efficient interaction and understanding of shared artifacts and objects with excellent operability

**Table 7** Collaboration awareness scales

Level	Quality description
Low ( $\theta < -1$ ) ( $\theta^* < 90$ )	The collaborative environment rarely provides social interaction and collaboration aspects nor considers the participant's capabilities. Some basic identity and communication resources are reached. Due to these limitations, the collaboration aspects are limited
Good ( $-1 \leq \theta \leq 1$ ) ( $90 \leq \theta^* \leq 110$ )	The collaborative environment sometimes presents social interaction and collaboration aspects. The environment provides moderate status and identity information. Sometimes, it is considered the participant's capabilities. The awareness information is often considered relevant to the participant's interests, and they usually recognize that the content helps the collaboration process
Excellent ( $\theta > 1$ ) ( $\theta^* > 110$ )	At this level, the collaborative environment is challenging for group members and presents no difficulties for interaction. It is highly relevant to participants' interests and provides excellent focused attention and social interaction. In terms of usability, the environment presents excellent operability; it has clear rules and is easy to interact with

**Table 8** Contextual awareness scales

Level	Quality description
Low ( $\theta < -1$ ) ( $\theta^* < 90$ )	The collaborative environment hardly considers the contextual perspective nor the group members' mobility. Navigation or spatiality is rarely allowed, and contextual interaction is limited
Good ( $-1 \leq \theta \leq 1$ ) ( $90 \leq \theta^* \leq 110$ )	The collaborative environment provides moderated access to contextual information. Participants partially reach environmental navigation and spatiality aspects. The environment provides some operability over participants' contextual information; however, members' mobility remains not archived
Excellent ( $\theta > 1$ ) ( $\theta^* > 110$ )	At this level, the collaborative environment provides clear access to contextual information. Participants reach environmental navigation and spatiality aspects, and the environment presents certain operability over participants' contextual information

in which participants present a certain probability  $P_i$  of selecting each response category presented.

In other words, starting from the participants' ability scale  $\theta$ , the model represents the probable intervals  $P_i(\theta)$  that participants are most likely to correctly identify/understand the awareness mechanism in the evaluated interface.

For example, as demonstrated in 9c, participants with ability score  $\theta \leq 0$  answered "strongly disagree" or "disagree" for all contextual awareness mechanisms, which indicates that contextual elements are hard to identify in the evaluated environments, or even, that they require a higher level participant skill/expertise. Only participants with an ability level  $\theta \geq 0$  identify these elements and only

participants with an ability level  $\theta \geq 2$  (experts) strongly agreed with these mechanisms.

In the second awareness scale perspective, presented in Tables 6, 7, and 8, we categorized the results concerning the skill levels of the expected participants. In the assessment model, the ability scale  $\theta$  encompasses within the interval  $[-4.0, +4.0]$  and the adjusted ability scale  $\theta^*$  encompasses within the interval  $[+60, +140]$  (awareness points). Then, the assessment scale established three participant ability intervals, describing the expected competencies concerning the awareness mechanisms participants in each ability score interval understand.

In the workspace, collaboration, and awareness quality scale, the awareness mechanisms are organized in a

gradual acquisition perspective, indicating which awareness mechanisms are supported/understood by novices, intermediates, and expert participants. This gradual organization allows us to prioritize mechanisms from participants' ability perspective, providing insights regarding adjustments and/or necessary modifications to enable participants with lower ability skills (novices) to easily acquire the more important awareness mechanisms.

### 7 Model reliability

The reliability of a set of items is one of the properties to evaluate the quality of the instrument. Similarly to classical approaches, one of the ways to check internal consistency is through Cronbach's alpha coefficient ( $\alpha$ ) [25]. This coefficient is calculated based on the values obtained through the application of data collection and analysis tools. Cronbach's alpha coefficient indicates the degree to which a set of items measure a single factor [20]. In addition, IRT allows us to evaluate the assessment items' quality through  $\theta$ , discrimination, and difficulty parameters.

Cronbach's alpha can be written as a function of the number of test items, and the average inter-correlation among items is given by Eq. 6, where  $N$  represents the number of items,  $\bar{c}$  is the average inter-item covariance among the items, and  $\bar{v}$  equals the average variance. We consider values of Cronbach's alpha between  $0.8 > \alpha \geq 0.7$  acceptable; between  $0.9 > \alpha \geq 0.8$  good; and  $\alpha \geq 0.9$  excellent [25].

$$\alpha = \frac{N\bar{c}}{\bar{v} + (N - 1)\bar{c}} \tag{6}$$

where  $\alpha$  is the Cronbach's alpha coefficient;  $N$  represents the number of items;  $\bar{c}$  is the average inter-item covariance;  $\bar{v}$  equals the average variance.

The general quality of a collaborative environment is determined based on the data collected using the measurement instrument and analyzing them through the ability level ( $\theta$ )'s scale scores. To assess reliability, the assessment model uses the IRT technical properties of discrimination ( $a$ ) and difficulty ( $b$ ), combined with  $\alpha$  coefficient, to assess the Awareness Assessment Model Instrument reliability and internal consistency.

#### 7.1 Reliability results

Both alpha and IRT params strongly demonstrate the validation of the proposed model. First, the adequate representation of the awareness scale  $\theta$  (interval  $[-2.96, +2.70]$ ) as

presented in Fig. 8) is good evidence of the instrument's reliability. In addition, the internal reliability through Cronbach's alpha coefficient [25] demonstrated an excellent internal consistency for all assessment items ( $\alpha > 0.91$ ).

Second, we calculate the instrument's reliability function  $r_{xx}(\theta)$  [18, 96], across participants' latent trait (see Fig. 10). The model shows excellent reliability, and the function reaches its highest value ( $r_{xx} > 0.90$ ) over the scale region where the information function is representative.

### 8 Model dimensionality

An essential factor that corroborates the IRT model's validation is the latent trait's dimensionality, which, in this case, refers to the number of factors necessary to explain the variability of the data and constitute a hypothesis to be verified [86]. IRT models can result in a unidimensional character when there is only one factor under analysis or multidimensional when there is more than one determining factor. There must be a single ability responsible for performing all test items.

To satisfy the unidimensionality postulate, it is sufficient to admit that a dominant ability is being measured (a dominant factor) and responsible for the set of items [22]. This factor is what is supposed to be measured by the assessment instrument. Schmitt [84] emphasizes that the more strictly unidimensional the construct, the less ambiguous its interpretations become, and consequently, its correlations become more legitimate.

Therefore, dimensionality is an intrinsic factor to the construct and defines the homogeneity of the set of items. Disregarding this factor results in an improperly applied measurement model, generating erroneous inferences about the evaluation of results and may threaten the credibility of the measurement instrument [87].

We used Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) to test dimensionality. EFA aims to identify the underlying relationships between the measured items and evaluate the dimensionality of a series of items to identify the smallest number of latent traits that

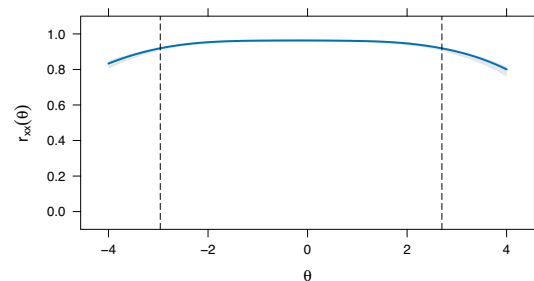


Fig. 10 Test reliability

explain the correlations pattern [69]. Confirmatory factor analysis (CFA) is used to verify the factor structure of a set of observed variables. It allows us to test the hypothesis that a relationship exists between observed variables and their underlying latent constructs [17, 97].

Due to the sample size, the same data set was used for the EFA and CFA. In this configuration, Izquierdo et al. [50] highlights that CFA results provide good fit indices and conform to the scale structure discovered in EFA as they were calculated based on the same data.

To determine the number of factors retained in the EFA, we will use the Latent Root (or Kaiser) Criterion [45]. In the Latent Root, the factors or components retained in the analysis with real data must have an eigenvalue higher than ones obtained randomly [57]; thus, only factors with eigenvalues greater than or equal to 1 are considered.

In general terms, factor analysis addresses the problem of analyzing the structure of interrelationships (correlations) between a large number of variables (e.g., test scores, test items, questionnaire responses) by defining a set of latent dimensions, called factors [45].

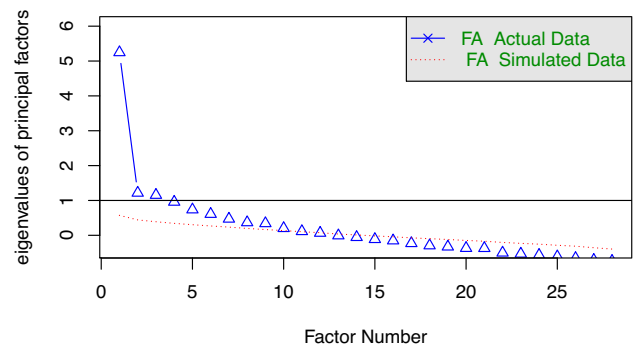
Generally, the eigenvalues of the correlation matrix or covariance matrix are used to decide the number of factors to be extracted. Factor loadings equal to or greater than 0.30 are considered high factor loading for samples larger than 350 observations [71]. In contrast, items with factor loading below would not measure the same thing as the others, i.e., do not have a large enough charge to merit interpretation [71]. This technique allows data reduction by eliminating variables with little loading, identifying the most representative variables, or creating a new set of variables much smaller than the original [45].

## 8.1 EFA results

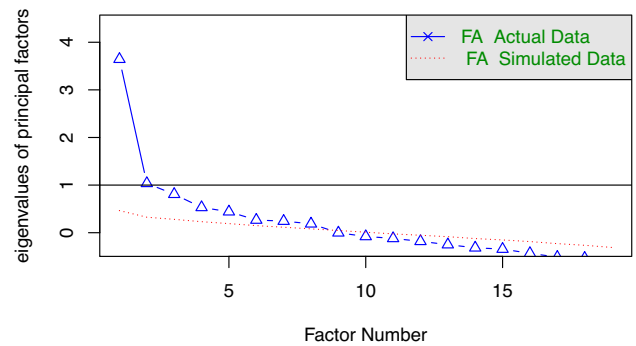
Factor analysis is based on the simulation of random data to determine the number of factors [50]. The factors/components retained with real data must have an eigenvalue higher than those obtained randomly.

As shown in Fig. 11, the latent root criterion suggests a strong principal component, with three other prominent components (11a–c). By applying the Kaiser criterion (eigenvalue  $\geq 1$ ), we identify three principal components for the workspace, two for the collaboration and two for the contextual category.

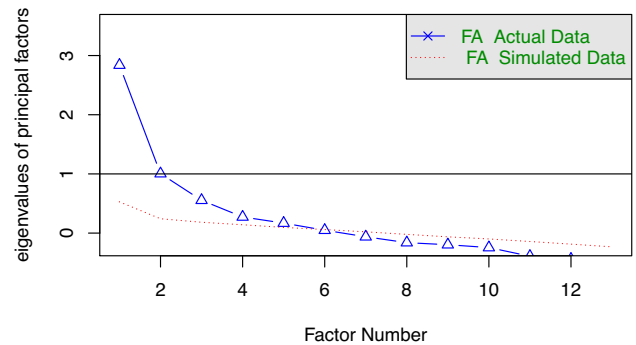
The inclination angles decreased sharply from the second factor onwards, approaching the horizontal line of value one and converging to the red dotted line of the EFA simulated data. These characteristics corroborate a representative model for each awareness category. Despite slightly indicating a secondary component, we have verified the simpler and highly representative IRT unidimensional model.



(a) Workspace awareness



(b) Collaboration awareness



(c) Contextual awareness

Fig. 11 Model dimensionality

The explanation power of the factors relative to the total variance is explained as follows. In the workspace awareness perspective, Factor 1 explains 21.49% ( $pc = 6, 0161$ ); in the contextual awareness perspective, Factor 1 explains 21.83% ( $pc = 4, 3831$ ); and in the collaboration awareness perspective, Factor 1 explains 23.07% ( $pc = 3, 5431$ ) of the total variance. Literature suggests that factor analysis results may indicate unidimensionality if the first factor is greater than or equal to 20% of the total eigenvalue of the principal components variance [76].



### 8.2 CFA results

In the EFA analysis, we combined the conceptual rationale with the empirical evidence extracted from the model to identify the underlying relationships between measured items and the smallest number of latent traits that explain the pattern of correlations [45]. The awareness perspectives were rewritten considering three factors for the workspace, two for the collaboration, and two for the contextual awareness perspectives.

Figure 12 contains the graphical representation of the model, correlating the factors and their related awareness

elements (assessment items) and the factor loadings. A factor loading of more than 0.30 usually indicates a moderate correlation between the item and the factor [71, 94].

Extracting the evidence from the factor analysis of the model presented in Fig. 12, we identified the main latent dimensions (factors) for the workspace and two for collaboration and contextual awareness perspective, as shown in Tables 9, 10, and 11.

Comparing the model generated by the EFA analysis (Fig. 12 the assessment items of the taxonomy (awareness elements and design categories), we could visualize a significant equivalence. Furthermore, as shown in Fig. 12, all

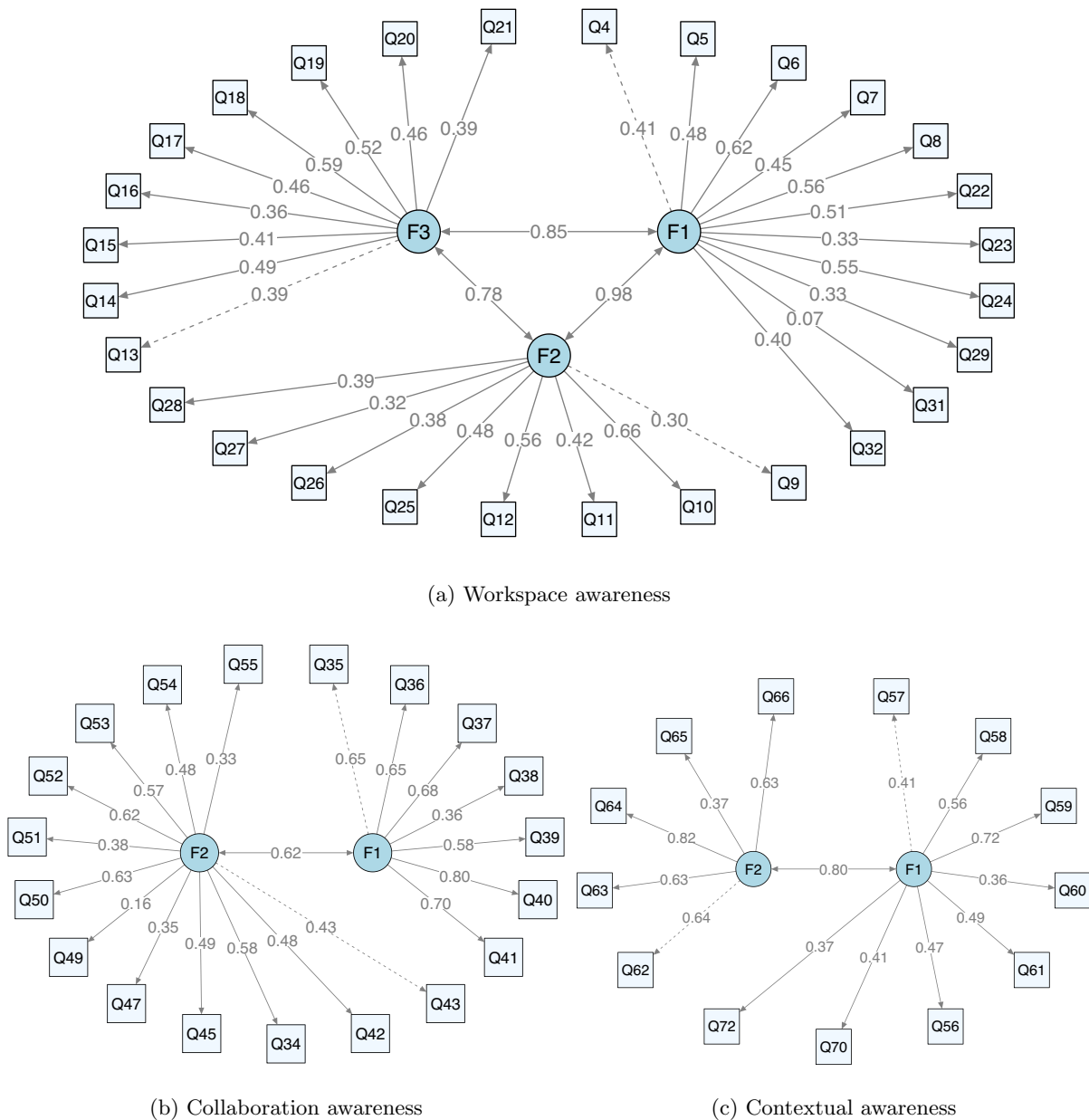


Fig. 12 EFA (estimated model)

**Table 9** Factor analysis (workspace awareness)

Factor	Factor description
F1	Factor 1 provides an understanding of the collaborative activities; it involves the assessment items Q4 to Q8 (motivation, time required, progress level, help needed, and evaluation), Q22 to Q24 (meaning, scenarios, and sense-making), Q29 (action control), Q31 (access privileges), and Q32 (control mechanisms)
F2	Factor 2 represents the participant's feedback over the shared workspace; it involves the assessment items Q9 to Q12 (authorship, execution steps, events and actions, and change locations) and Q25 to Q28 (feedback, feedthrough, backchannel feedback, and feed-forward feedback)
F3	Factor 3 represents the resources available in the shared workspace; it involves the assessment items Q13 to Q21 (related, parallel, coordinated, and mutually adjusted activities, tools and materials, artifacts and objects, resources availability, critical elements, and virtual relationship)

**Table 10** Factor analysis (collaboration awareness)

Factor	Factor description
F1	Factor 1 represents the participant's capabilities to collaborate; it refers to the assessment items Q35 to Q41 (preferences, rules, responsibilities, privileges, knowledge, influences, and intentions)
F2	Factor 2 represents the resources to establish communication among participants; it refers to the assessment items Q34 (shared profile), Q42 (availability), Q43 (presence), Q45 (status), Q47 (connectivity), and Q49 to Q55 (message delays, interaction ways, turn-taking, conversation, expectations, emotional status, and non-verbal cues)

**Table 11** Factor analysis (contextual awareness)

Factor	Factor description
F1	Factor 1 represent a virtual setting of the contextual environment; it involves the assessment items Q56 to Q61 (location, distances, restrictions, places, topology, and attributes), Q70 (voice cues), and Q72 (eye-gaze cues)
F2	Factor 2 represent common information about the spatiality of the shared environment; it involves the assessment items Q62 to Q66 (spatiality attributes like view, reach, orientation, movement, and range of attention)

factor loadings exceed 0.3. The confirmatory factor analysis results largely maintained the taxonomy's structure, and the CFA items' factor loadings demonstrate the instrument's construct validity.

Finally, Composite Reliability (CR) was calculated to evaluate the construct validity of the proposed model [45]. The measurements obtained were then evaluated following the Fornell and Larcke recommendations [33]. All factors in

the model must present a CR value above 0.7 to demonstrate the instrument's construct validity [33, 45].

Evidence of construct validity indicates that the items measured in the sample represent the real measurements in the population [45].

In the contextual awareness perspective, factors F1, F2, and F3 presented CR values of 0.712, 0.692 ( $\approx 0.7$ ), and 0.706. In the collaboration awareness perspective, factors F1 and F2 presented CR values of 0.825 and 0.761, respectively. In the contextual awareness perspective, values of 0.708 and 0.715 were found for factors F1 and F2.

## 9 Discussion

Regarding the case study, we obtained voluntary participation from 422 individuals who answered one of the ten questionnaires (test books) provided in the full version of the model. As a result, we found suitable indicators from the perspective of demographic data and IRT parameterization. Then, skill and awareness quality scales were constructed based on the 60 calibrated items.

The results of the videoconferencing assessment were positive, and the most familiar environments presented the best performance. Moodle (Big Button Blue), Google Meet, and Microsoft Teams were the environments that presented a lower overall awareness score, respectively,  $\theta$  equals to  $-0.21$ ,  $-0.12$ , and  $0.12$ ; the adjusted ability scores  $\theta^*$  were equal to 97.9, 98.8, and 101.2 awareness points. Users of Zoom, Skype, and Discord indicated a slightly greater facility in identifying awareness information, respectively,  $\theta$  equals to 0.16, 0.20, and 0.21; the adjusted ability scores  $\theta^*$  were equal to 101.6, 102.0, and 102.1 awareness points.

Our awareness quality scale was established considering the participants' ability to identify awareness information; consequently, higher scores indicate that evaluated environments easily support awareness mechanisms, whereas participants with higher ability scores can identify properly existing awareness mechanisms.

Estimating IRT parameters with a low standard error and positioning items on the scale requires many respondents per item category. Few items did not present an ideal calibration and were excluded from the interpretation scale phase due to an outlier of  $\theta$ ,  $a$ , or  $b$ . In these cases, there was no adequate variability in the responses obtained (strongly disagree to strongly agree), making a fair analysis impossible.

We carefully evaluated the items and come to the conclusion that non-calibration occurred due to many positive (agree or strongly agree) or negative (disagree or strongly disagree) responses. In the first, the analysis suggests that most participants found it easy to identify the awareness mechanism and judge the assessment item; in the second, the participants had difficulty identifying the element, or this aspect was absent in the evaluated environment.

To construct the ability scale  $\theta$ , the assessment model calculates the probability  $P_{i,k}(\theta_j)$  considering the gradual scale of [81]. The generated awareness support scale presented a coverage interval  $[-4.0, +4.0]$ , with the most appropriate values in interval  $[-2.96, +2.70]$ . Although there are no practical differences in establishing the first or the second one, we minimized the eventual negative impact or misinterpretation of a participant with a low ability score represented with negative values in the final scale by applying a linear conversion and generating a positive scale.

The literature review found no references regarding a scale for assessing awareness using Item Response Theory. In addition, the awareness scale presented has several levels of the latent trait that make it possible to interpret the degree of skill that an evaluator has given that he or she used the proposed measurement instrument.

Validation of the model through the expert panel and case study approaches was very positive. In the first, we analyze the usefulness aspects, namely, clarity, relevance, consistency, and completeness of the measurement instrument, resulting in a refined version of the artifacts; in the second, we exposed the model in a videoconferencing assessment scenario to assess the model's internal consistency, reliability, and dimensionality.

In all 60 calibrated assessment items,  $\theta$ ,  $a$ , and  $b$ , combined with the internal consistency values of Cronbach's alpha and reliability function  $r_{xx}(\theta)$ , indicates an excellent instrument's reliability.

We used exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) to test the model dimensionality. The EFA results indicated a strong tendency towards the one-dimensional model (latent root criterion [45]), legitimating the correlation between the assessment items and the observed latent trait. The CFA results demonstrate the instrument construct validity: all factors presented adequate composite reliability ( $CR \geq 0.7$ ) [33, 45], and all factor

loadings of the assessment items are above the limit ( $\geq 0.3$ ) [71, 94].

## 9.1 Model limitations

An initial limitation is related to the number of respondents. IRT requires very large sample sizes for many models, often exceeding what is typically used in classical theory research. According to IRT, the sample size to perform an item analysis depends on the number of model parameters and item categories; in other words, it depends on the number of parameters to be estimated.

Our study obtained 422 observations to estimate IRT params  $a$ ,  $b$ , and  $\theta$  of a universe of 60 calibrated assessment items, each with four response categories (Samejima's gradual scale [81]). Thus, we obtained ten or more observations for each response category; where this criterion was not met, we grouped the response categories (strongly disagree and disagree or agree and strongly agree). This grouping may eventually affect the item calibration.

The IRT model could not calibrate some items from the 75 assessment items originally proposed in the taxonomy, as discussed in section 6.1. Thus, validating these items was also impossible in this work; alternatively, we suggest new evaluation scenarios to investigate their suitability for accessing awareness support in collaborative environments.

Regarding the number of evaluation items and the sample size (observations) used in the evaluated scenario, we can present two main aspects discussed in the following.

**Assessment items.** Although the assessment model presents in its conceptual view (awareness taxonomy) 75 assessment items/awareness mechanisms, we provide for both the assessment process and the assessment protocol artifacts the possibility for the examiner to choose which awareness categories and awareness items will be used during the evaluation process of the collaborative environment. Thus, the model can be adjusted to direct an assessment compatible/appropriate to the use context. This paper seeks to evaluate as many assessment items as possible. Due to this, the full version of the 75 items divided into ten balanced test notebooks (blocks) was adopted for evaluation.

We used the BIB (Balanced Incomplete blocks) strategy to create the ten notebooks with a replication factor ( $r = 4$ ). A Balanced Incomplete Block (BIB) consists of treatments  $t$  (a subset of the assessment items) that appear in the same block  $b$  (questionnaire) with each other treatments the same number of times  $\lambda$ . Thus, each notebook is represented by two treatments with different assessment items ( $t = 15$ ). Therefore, in the ten test books (questionnaires) identified, we have the repetition of four combinations of treatments (see Table A1).

The adoption of BIB did not happen arbitrarily. The BIB strategy, combined with the statistical model of Item Response Theory, allows the ability scale ( $\theta$ ) to be on the same scale for all test books; thus, each test book is equally representative in its own right. After building and calibrating the ITR model, if we adopt one of the questionnaires, apply all ten test books, or even a subset of these, we find the same skill scale; therefore, the number of notebooks used is irrelevant. In other words, the generated scale will be equivalent if we choose just one of the ten bib blocks after model calibration.

Some preliminary applications that we are developing demonstrate this possibility, reducing the items and the required sample size. It is also important to note that the application context alone is also an indicator that will indicate the need to remove some items.

**Sample size.** Despite the large number of items (75), it was possible to calibrate most (60). Uncalibrated items refer to participants' unanimous positive responses. In this case, the statistical method (TRI) suggests that such items should be ignored, as they are not representative for the discrimination parameters ( $a$ ). Although they are important awareness mechanisms, they are insignificant for positioning and generating the awareness skill scale ( $\theta$ ).

The main objective of eliminating items during the scale construction is to simplify the assessment model; retaining only those items with an acceptable discrimination value ( $> 0.65$ ). Furthermore, for different applications, a subset of the model can be used, adopting the parameters  $a$  and  $b$  calibrated in this paper—which considerably simplifies the required number of evaluation items and required sample size.

Finally, although a relatively small sample was used, the IRT model converged to notable parameters due to sampling variability across the skill spectrum ( $\theta$  in the range between  $[-4, +4]$ ); as seen in Figs. 6 and 7, both demographic distribution and normal and accumulated distribution corroborate the thesis of a model strongly representative of the assessed latent trait (the conscious support).

Another limitation is the complexity and difficulty of performing IRT analyses. These analyses require specialized knowledge to perform tests of assumptions and estimation of parameters and tests for model adjustment. In this sense, we carefully designed calibration and estimation scripts for the IRT model and provided all the necessary artifacts for using the model available in the repository [60].

## 10 Conclusions and future work

The awareness and collaboration concepts are intrinsically related since the foundation and their understanding expanded in the same way as research in the field evolves.

We observed the efforts towards establishing common sense about what awareness is, what it represents, and what it is related to. On the other hand, achieving an accurate and clear-cut definition of awareness remains challenging.

Understanding and providing aspects of collaboration likewise involves a comprehensive knowledge of the elements of awareness that support it. We do not envision ways to provide efficient communication, coordination, or cooperation without proper awareness support. Provisioning adequate awareness mechanisms ensures the support for the whole collaboration process, consolidating awareness as the cornerstone of collaborative environments.

This work applies efforts toward building a model for evaluating awareness support in collaborative environments from the users' point of view. Assuming a plural collaborative environment, where different participants with different skills, knowledge, and wisdom meet and interact, the model seeks to build a more faithful representation of these existing profiles across a broad spectrum of individual abilities.

We present a new assessment method for awareness and collaboration support centered on the participant's perspective by developing a measurement instrument based on Item Response Theory. The methodology allowed us to construct and interpret an awareness quality scale to evaluate the support level for three awareness dimensions and 75 assessment items. Consequently, we argue that the essential aspects of the collaboration process are provided through adequate support for each awareness view. The correlations between design and awareness elements were defined according to theory and practice.

The method can be replicated by applying the artifacts described in the model (available at model dataset [60]). To use the proposed assessment method properly, this model includes an assessment process inspired by the ICH guidelines and the recommendations for evaluating software product quality of ISO/IEC 25040:2011 [32]. In this way, an adaptive approach was designed, where the examiner can apply the complete assessment model or select the respective design categories and assessment elements of interest, thus adjusting the data collection and analysis artifacts. With IRT, it is possible to include new items in the same measurement scale, like demographic, usability, and UX, increasing the evaluation potential.

The statistical method for composing our assessment model, IRT, notably involves heavy calculations, which were the focus of this paper. All the steps described aim for examiners to understand all the steps of statistical calculations used (which are the basis of IRT), like the processing of raw data/observations collected, calibration, construction of support scales, and the validation of the model by different necessary statistical approaches (Cronbach's alpha coefficient, EFA, and CFA). Describing all these processing steps enables the model's reproducibility

through the data available in our awareness assessment model repository.

We believe that awareness is intrinsically linked to the participant's skills in identifying, understanding, and projecting their actions. Thus, properly assessing a collaborative environment support is possible if the assessment considers the awareness elements from the participant's perspective. As participants' understanding differs, the awareness support scale must represent individuals with lower or higher abilities.

This article focused on the evaluation model itself. Indications for development or other stages of building a collaborative and/or general-purpose application project were not considered—although the artifacts contained in the conceptual view of the model can be used as a set of recommendations that developers can use, for example. From this perspective, the design categories and awareness mechanisms, especially in the awareness taxonomy, can be used as potential requirements for collaborative applications to support awareness mechanisms adequately.

Designing collaborative applications must consider an appropriate set of awareness mechanisms, which largely depend on the application's objectives, context, target users, etc. Any recommendations for using the awareness mechanisms presented in this paper must be evaluated based on their suitability for other scenarios; a suitable set of awareness mechanisms for a given context may not be the same for another. For this reason, the assessment model specification allows the awareness support scale to be created for each item in the model, category, dimension of awareness, or even other specific subset.

In the awareness assessment repository [60], all model artifacts, both the conceptual view and the assessment process artifacts, are available. The latter presents a guide/protocol for evaluating awareness support through the IRT model. We designed a detailed set of artifacts to help use the model.

In future work, we suggest validating the assessment model in other collaborative environments and contexts to verify its flexibility and other examiners' replicability of the assessment process.

As some of the 75 initial assessment items were not calibrated properly in the IRT model of the proposed case study, we encourage new scenarios to verify the items' behavior in other contexts.

In the proposed case study scenario, the full version of the assessment model was applied; thus, new studies may be necessary to verify the model's applicability by considering a subset of the assessment items.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10209-024-01110-5>.

**Availability of data and materials** All the steps in this research, including the planning, extraction, analysis, and synthesis, were documented, and the dataset is available at [60].

## Declarations

**Conflict of interest** The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

**Authors' contributions** All authors contributed to the study's conception and design, including the material preparation, data collection, and analysis. The manuscript was written by MJM and revised by FBVB. All authors contributed to the manuscript revision and read and approved the submitted version.

## References

1. Agogo, D., Hess, T.J.: "How does tech make you feel?" A review and examination of negative affective responses to technology use. *Eur. J. Inf. Syst.* **27**(5), 570–599 (2018). <https://doi.org/10.1080/0960085X.2018.1435230>
2. Altenburger, T., Guerriero, A., Vagner, A., et al.: Groupware requirements modelling for adaptive user interface design. In: *eWork and eBusiness in Architecture, Engineering and Construction—Proceedings of the European Conference on Product and Process Modelling 2012, ECPPM 2012*, vol. 1, no. 1, pp. 825–832 (2012). <https://doi.org/10.1201/b12516-131>
3. Antunes, P., Sapateiro, C., Pino, J.A., et al.: Awareness checklist: reviewing the quality of awareness support in collaborative applications. *Collab. Technol.* **1**(1), 202–217 (2010). [https://doi.org/10.1007/978-3-642-15714-1\\_16](https://doi.org/10.1007/978-3-642-15714-1_16)
4. Antunes, P., Herskovic, V., Ochoa, S.F., et al.: Structuring dimensions for collaborative systems evaluation. *ACM Comput. Surv.* **44**(2), 1–28 (2012)
5. Antunes, P., Herskovic, V., Ochoa, S.F., et al.: Reviewing the quality of awareness support in collaborative applications. *J. Syst. Softw.* **89**(1), 146–169 (2014). <https://doi.org/10.1016/j.jss.2013.11.1078>
6. Bailey, K.D.: *Typologies and taxonomies: An introduction to classification techniques*, vol. 102. Sage, London (1994)
7. Baker, F.B.: *The Basics of Item Response Theory*. ERIC, Washington (2001)
8. Baker, F.B., Kim, S.H.: *Item Response Theory: Parameter Estimation Techniques*. CRC Press, Boca Raton (2004)
9. Baker, F.B., Kim, S.H.: *The Basics of Item Response Theory Using R*. Springer, New York (2017). <https://doi.org/10.1007/978-3-319-54205-8>
10. Barbosa, S., Silva, B.: *Interação humano-computador*. Elsevier, Brasil (2010)
11. Basili, V.: Software modeling and measurement: The goal/question/metric paradigm. Technical Report (1992)
12. Beecham, S., Hall, T., Britton, C., et al.: Using an expert panel to validate a requirements process improvement model. *J. Syst. Softw.* **76**(3), 251–275 (2005). <https://doi.org/10.1016/j.jss.2004.06.004>
13. Belkadi, F., Bonjour, E., Camargo, M., et al.: A situation model to support awareness in collaborative design. *Int. J. Hum. Comput. Stud.* **71**(1), 110–129 (2013). <https://doi.org/10.1016/j.ijhcs.2012.03.002>
14. Bodemer, D., Dehler, J.: Group awareness in CSCL environments. *Comput. Hum. Behav.* **27**(3), 1043–1045 (2011). <https://doi.org/10.1016/j.chb.2011.03.002>

- [doi.org/10.1016/j.chb.2010.07.014](https://doi.org/10.1016/j.chb.2010.07.014). (group Awareness in CSCL Environments)
15. Bødker, S., Christiansen, E.: Computer support for social awareness in flexible work. *Comput. Support. Coop. Work* **15**(1), 1–28 (2006). <https://doi.org/10.1007/s10606-005-9011-y>
  16. Brezillon, P., Borges, M., Pino, J., et al.: Context-based awareness in group work. In: *Proceeding of the 17th International FLAIRS Conference*. AAAI Press, Miami (2004)
  17. Brown, T.A.: *Confirmatory Factor Analysis for Applied Research*. Guilford Publications, New York (2015)
  18. Chalmers, R.P.: mirt: a multidimensional item response theory package for the R environment. *J. Stat. Softw.* (2012). <https://doi.org/10.18637/jss.v048.i06>
  19. Collazos, C.A., Gutierrez, F.L., Gallardo, J., et al.: Descriptive theory of awareness for groupware development. *J. Ambient. Intell. Humaniz. Comput.* **10**(12), 4789–4818 (2019)
  20. Cronbach, L.J.: Coefficient alpha and the internal structure of tests. *Psychometrika* **16**(3), 297–334 (1951)
  21. Cruz, A., Correia, A., Paredes, H. et al.: Towards an overarching classification model of cscw and groupware: a socio-technical perspective. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Raesfeld, Germany, pp. 41–56 (2012). [https://doi.org/10.1007/978-3-642-33284-5\\_4](https://doi.org/10.1007/978-3-642-33284-5_4)
  22. de Andrade, D.F., Tavares, H.R., da Cunha, V.R.: Teoria da resposta ao item: conceitos e aplicações. ABE, Sao Paulo (2000)
  23. De Araújo, L.P., Mantau, M.J., Citadin, J.R., et al.: Heuristic evaluation for mobile groupware: evaluating two audience response systems. *J. Appl. Comput. Res.* **3**(2), 64–77 (2014)
  24. De Souza, L.G., Barbosa, S.D.J.: Evaluating the molicc notation using the cognitive dimensions of notations framework. In: *Proceedings of the 14th Brazilian Symposium on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, IHC '15, pp. 1–10 (2015). <https://doi.org/10.1145/3148456.3148474>
  25. DeVellis, R.F.: *Scale Development: Theory and Applications*, Applied Social Research Methods, vol. 26, 4th edn. Sage Publications Inc, London (2016)
  26. Do Espírito Santo, D.A., De Oliveira Rodrigues, B.R., Hadad Zaidan, F., et al.: Quality assessment of awareness support in agile collaborative tools. In: *Proceedings—2018 44th Latin American Computing Conference, CLEI 2018*, Sao Paulo, Brazil, pp. 21–30 (2018). <https://doi.org/10.1109/clei.2018.00013>
  27. Dourish, P., Bellotti, V.: Awareness and coordination in shared workspaces. In: *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work*. Association for Computing Machinery, New York, NY, USA, CSCW '92, pp. 107–114 (1992). <https://doi.org/10.1145/143457.143468>
  28. Dourish, P.: *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press, Massachusetts Institute of Technology, Ohio, USA (2001). <https://doi.org/10.7551/mitpress/7221.001.0001>
  29. El Helou, S., Tzagarakis, M., Gillet, D., et al.: Participatory design for awareness features: Enhancing interaction in communities of practice. In: *Proceedings of the 6th International Conference on Networked Learning*, vol 1, no. 1, pp 23–530 (2008)
  30. Ellis, C., Gibbs, S., Rein, G.: Groupware some issues and experiences. *Commun. ACM* **34**(1), 38–58 (1991). <https://doi.org/10.1145/99977.99987>
  31. Engelmann, T., Dehler, J., Bodemer, D., et al.: Knowledge awareness in CSCL: a psychological perspective. *Comput. Hum. Behav.* **25**(4), 949–960 (2009). <https://doi.org/10.1016/j.chb.2009.04.004>. (including the Special Issue: The Use of Support Devices in Electronic Learning Environments)
  32. for Standardization IO.: *Systems and software engineering—systems and software quality requirements and evaluation (square)—evaluation process*. Standard, International Organization for Standardization, Geneva, CH (2011)
  33. Fornell, C., Larcker, D.F.: Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* **18**(1), 39–50 (1981)
  34. Gallardo, J., Molina, A.I., Bravo, C., et al.: An ontological conceptualization approach for awareness in domain-independent collaborative modeling systems: application to a model-driven development method. In: *Expert Systems with Applications*, pp. 1099–1118 (2011). <https://doi.org/10.1016/j.eswa.2010.05.005>
  35. Gallardo, J., Bravo, C., Molina, A.I.: A framework for the descriptive specification of awareness support in multimodal user interfaces for collaborative activities. *J. Multimodal User Interfaces* **12**(2), 145–159 (2018)
  36. Gerosa, M., Fuks, H., Lucena, C.: Suporte a percepção em ambientes digitais de aprendizagem. *Revista Brasileira de Informática na Educação. Soc. Bras. Comput.* **11**, 75–85 (2003)
  37. Gold, R., Mascolo, C.: Use of context-awareness in mobile peer-to-peer networks. In: *Proceedings Eighth IEEE Workshop on Future Trends of Distributed Computing Systems, FTDCS 2001*, pp. 142–147 (2001)
  38. Greenberg, S.: Collaborative interfaces for the web. *Hum. Factors Web Dev.* **18**, 241–254 (1997)
  39. Gross, T.: Supporting effortless coordination: 25 years of awareness research. *Comput. Support. Coop. Work* **22**(4–6), 425–474 (2013). <https://doi.org/10.1007/s10606-013-9190-x>
  40. Gutwin, C., Barjawi, M., Pinelle, D.: The emergence of high-speed interaction and coordination in a (formerly) turn-based groupware game. In: *Proceedings of the 19th International Conference on Supporting Group Work*. Association for Computing Machinery, New York, NY, USA, GROUP '16, pp. 277–286, (2016). <https://doi.org/10.1145/2957276.2957315>
  41. Gutwin, C., Bateman, S., Arora, G. et al.: Looking away and catching up: dealing with brief attentional disconnection in synchronous groupware. In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. Association for Computing Machinery, New York, NY, USA, CSCW '17, pp. 2221–2235 (2017). <https://doi.org/10.1145/2998181.2998226>
  42. Gutwin, C., Stark, G., Greenberg, S.: Support for workspace awareness in educational groupware. In: *The First International Conference on Computer Support for Collaborative Learning*. L. Erlbaum Associates Inc., USA, CSCL '95, pp. 147–156 (1995). <https://doi.org/10.3115/222020.222126>
  43. Gutwin, C., Greenberg, S.: A descriptive framework of workspace awareness for real-time groupware. *Comput. Support. Coop. Work* **11**(3–4), 411–446 (2002). <https://doi.org/10.1023/a:1021271517844>
  44. Haddaway, N.R., Page, M.J., Pritchard, C.C., et al.: Prisma 2020: an r package and shiny app for producing Prisma 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. *Campbell Syst. Rev.* **18**(2), e1230 (2022). <https://doi.org/10.1002/cl2.1230>
  45. Hair, J.F., Black, W.C., Babin, B.J. et al.: *Análise multivariada de dados*. Bookman editora (2009)
  46. Hakkila, J., Mantyjarvi, J.: Developing design guidelines for context-aware mobile applications. In: *Proceedings of the 3rd International Conference on Mobile technology, Applications and Systems*. ACM, New York (2006)
  47. Herskovic, V., Ochoa, S.F., Pino, J.A., et al.: The iceberg effect: behind the user interface of mobile collaborative systems. *JUCS J. Univ. Comput. Sci.* **17**(2), 183–201 (2011)
  48. Hinkelmann, K., Kempthorne, O.: *Design and Analysis of Experiments: Advanced Experimental Design*.

- Wiley-Interscience, New York (2005). <https://doi.org/10.1002/0471709948>
49. Hummel, D., Schacht, S., Maedche, A.: Determinants of multi-channel behavior: exploring avenues for future research in the services industry. In: ICIS (2016)
  50. Izquierdo, I., Olea, J., Abad, F.J.: Exploratory factor analysis in validation studies: uses and recommendations. *Psicothema* **26**(3), 395–400 (2014)
  51. Janssen, J., Erkens, G., Kirschner, P.A.: Group awareness tools: it's what you do with it that matters. *Comput. Hum. Behav.* **27**(3), 1046–1058 (2011). <https://doi.org/10.1016/j.chb.2010.06.002>
  52. Jonasson, N.: Designing for collaboration: interface guidelines for improving co-located collaboration on tabletop displays. In: Proceedings of Umea's 14th Student Conference in Computing Science—USCCS 2010, Umea University, Sweden, pp. 61–71 (2010)
  53. Khaliljafarabad, A., Helfert, M., Ge, M.: Developing a data quality research taxonomy—an organizational perspective. In: ICIQ, pp. 176–186 (2016)
  54. Kim, K.T., Javed, W., Williams, C., et al.: Hugin: a framework for awareness and coordination in mixed-presence collaborative information visualization. In: ACM International Conference on Interactive Tabletops and Surfaces, ITS, vol. 2010, pp. 231–240 (2010)
  55. Kitchenham, B., Charters, S.: Guidelines for performing systematic literature reviews in software engineering. University of Durham, Durham, UK, Technical Report (2007)
  56. Kreijns, K., Kirschner, P.A.: The social affordances of computer-supported collaborative learning environments. In: 31st Annual Frontiers in Education Conference. Impact on Engineering and Science Education. Conference Proceedings (Cat. No.01CH37193), FIE '01, vol 1. IEEE Computer Society, pp. T1F–12 (2001). <https://doi.org/10.1109/FIE.2001.963856>
  57. Ledesma, R.D., Valero-Mora, P.: Determining the number of factors to retain in EFA: an easy-to-use computer program for carrying out parallel analysis. *Pract. Assess. Res. Eval.* **12**(1), 2 (2019)
  58. Lopez, G., Guerrero, L.A.: Awareness supporting technologies used in collaborative systems: a systematic literature review. In: Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing. Association for Computing Machinery, New York, NY, USA, CSCW '17, pp. 808–820 (2017). <https://doi.org/10.1145/2998181.2998281>
  59. Mantau, M.J., Benitti, F.B.V.: Awareness support in collaborative system: reviewing last 10 years of CSCW research. In: 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD), pp. 564–569. IEEE (2022). <https://doi.org/10.1109/cscwd54268.2022.9776091>
  60. Mantau, M.J., Benitti, F.B.V.: The awareness assessment model repository. (2023). <https://doi.org/10.5281/zenodo.8298950>
  61. Mantau, M.J., Benitti, F.B.V.: Towards an awareness taxonomy. In: 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD), pp. 495–500. IEEE (2022). <https://doi.org/10.1109/cscwd54268.2022.9776129>
  62. Mantau, M.J., Berkenbrock, C.D.M., Berkenbrock, G.R.: Visualization and filtering awareness information in mobile groupware: an action research approach. In: Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD), pp. 563–568 (2014). <https://doi.org/10.1109/cscwd.2014.6846906>
  63. Mantau, M.J., Berkenbrock, C.D.M., Berkenbrock, G.R.: Análise de requisitos de percepção em um groupware móvel síncrono. *Rev. Sist. Inf. FSMA* **1**(19), 16–33 (2017)
  64. MI, B., Karahoca, D., Karahoca, A.: A measurement and structural model for usability evaluation of shared workspace groupware. *Int. J. Hum. Comput. Interact.* **34**(1), 35–56 (2018)
  65. Molina, A.I., Gallardo, J., Redondo, M.Á., et al.: Assessing the awareness mechanisms of a collaborative programming support system. *Dyna* **82**(193), 212–222 (2015). <https://doi.org/10.15446/dyna.v82n193.53497>
  66. Nickerson, R.C., Varshney, U., Muntermann, J.: A method for taxonomy development and its application in information systems. *Eur. J. Inf. Syst.* **22**(3), 336–359 (2013). <https://doi.org/10.1057/ejis.2012.26>
  67. Niemantsverdriet, K., Essen, H.V., Pakanen, M., et al.: Designing for awareness in interactions with shared systems: the DASS framework. *ACM Trans. Comput. Hum. Interact.* **26**(6), 1–41 (2019). <https://doi.org/10.1145/3338845>
  68. Nunnally, J.C., Bernstein, I.: *Psychometric Theory*, 3rd edn. McGraw-Hill Series in Psychology, McGraw Hill Higher Education, Maidenhead, England (1993)
  69. Osborne, J.W.: *Best practices in exploratory factor analysis*. CreateSpace Independent Publishing Platform (2014)
  70. Pasquali, L.: *TRI-Teoria de resposta ao item: Teoria, procedimentos e aplicacoes*. Editora Appris (2020)
  71. Pasquali, L., Primi, R.: Fundamentos da teoria da resposta ao item: Tri. *Aval. Psicol. Interam. J. Psychol. Assessment* **2**(2), 99–110 (2003)
  72. Petersen, K., Vakkalanka, S., Kuzniarz, L.: Guidelines for conducting systematic mapping studies in software engineering: An update. *Inf. Softw. Technol.* **64**, 1–18 (2015). <https://doi.org/10.1016/j.infsof.2015.03.007>
  73. Petticrew, M., Roberts, H.: *Systematic Reviews in the Social Sciences: A Practical Guide*. Blackwell Publishing Ltd, Oxford (2006). <https://doi.org/10.1002/9780470754887>
  74. Poulouvasilis, A., Xhafa, F., O'Hagan, T.: Event-based awareness services for p2p groupware systems. *Informatica (Netherlands)* **26**(1), 135–157 (2015). <https://doi.org/10.15388/informatica.2015.42>
  75. Prouzeau, A., Bezerianos, A., Chapuis, O.: Awareness techniques to aid transitions between personal and shared workspaces in multi-display environments. In: ISS 2018—Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces, Tokyo, Japan, pp. 291–304 (2018). <https://doi.org/10.1145/3279778.3279780>
  76. Reckase, M.D.: Unifactor latent trait models applied to multifactor tests: results and implications. *J. Educ. Stat.* **4**(3), 207–230 (1979). <https://doi.org/10.2307/1164671>
  77. Reinhardt, W., Mletzko, C., Sloep, P.B., et al.: Understanding the meaning of awareness in research networks. In: *CEUR Workshop Proceedings*, pp. 13–29. Saarbrücken, Germany (2012)
  78. Richardson, R.J.: *Pesquisa social: Métodos e técnicas*. Editora Atlas, S.A. (2017)
  79. Rocker, C.: Universal access to awareness information: using smart artefacts to mediate awareness in distributed teams. *Univ. Access Inf. Soc.* **11**(3), 259–271 (2012)
  80. Rogers, Y., Sharp, H., Preece, J.: *Design de interação*. Bookman Editora (2013)
  81. Samejima, F.: Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement* (1969). <https://doi.org/10.1007/bf03372160>
  82. Santos, V., Tedesco, P., Salgado, A.: Percepção e contexto. In: Pimentel, M., Fuks, H. (eds.) *Sistemas Colaborativos*, vol. 1, 1st edn., pp. 157–172. Elsevier, São Paulo (2011). (**chap 10**)
  83. Schmidt, K., Randall, D.: Preface to the special issue on 'reconsidering "awareness" in cscw'. *Comput. Support. Coop. Work (CSCW)* **25**(4), 229–233 (2016). <https://doi.org/10.1007/s10606-016-9257-6>

84. Schmitt, N.: Uses and abuses of coefficient alpha. *Psychol. Assess.* **8**(4), 350 (1996). <https://doi.org/10.1037/1040-3590.8.4.350>
85. Seebach, C., Beck, R., Pahlke, I.: Situation awareness through social collaboration platforms in distributed work environments. In: *International Conference on Information Systems 2011, ICIS 2011*, pp. 1–22. Shanghai, China (2011)
86. Singh, J.: Tackling measurement problems with item response theory: principles, characteristics, and assessment, with an illustrative example. *J. Bus. Res.* **57**(2), 184–208 (2004)
87. Spencer, S.G.: The strength of multidimensional item response theory in exploring construct space that is multidimensional and correlated. Brigham Young University (2004)
88. Sternberg, R., Mio, J.: *Cognitive Psychology*, 5th edn. Cengage Learning, Belmont (2006)
89. Sternberg, R.J., Sternberg, K., Mio, J.: *Cognitive Psychology*. Wadsworth, Cengage Learning, Belmont (2012)
90. Strasser, A.: Delphi method variants in information systems research: taxonomy development and application. *Electron. J. Bus. Res. Methods* **15**(2), 120–133 (2017)
91. Szopinski, D., Schoormann, T., Kundisch, D.: Because your taxonomy is worth it: towards a framework for taxonomy evaluation. In: *ECIS* (2019)
92. Talaei-Khoei, A., Vichitvanichphong, S., Solvoll, T., et al.: A methodology to develop awareness in computer supported collaborative work using policies. *J. Comput. Syst. Sci.* **80**(7), 1323–1338 (2014). <https://doi.org/10.1016/j.jcss.2014.03.003>
93. Tang, T.Y., Winoto, P., Leung, H.: A usability study of an educational groupware system: supporting awareness for collaboration. *J. Educ. Comput. Res.* **50**(3), 379–402 (2014). <https://doi.org/10.2190/ec.50.3.e>
94. Tavakol, M., Wetzell, A.: Factor analysis: a means for theory and instrument development in support of construct validity. *Int. J. Med. Educ.* **11**, 245–247 (2020). <https://doi.org/10.5116/ijme.5f96.0f4a>
95. Tenenberg, J., Roth, W.M., Socha, D.: From i-awareness to we-awareness in CSCW. *Comput. Support. Coop. Work (CSCW)* **25**(4), 235–278 (2016). <https://doi.org/10.1007/s10606-014-9215-0>
96. Thissen, D., Wainer, H.: *Test Scoring*. Routledge, London (2001)
97. Thompson, B.: Exploratory and confirmatory factor analysis: Understanding concepts and applications. In: Washington, DC, vol 10694, no. 3 (2004)
98. Tripathi, A.: A transaction model with multilevel consistency for shared data in distributed groupware systems. In: *Proceedings—2016 IEEE 2nd International Conference on Collaboration and Internet Computing, IEEE CIC 2016*, Pittsburgh, PA, United States, pp. 236–245. <https://doi.org/10.1109/cic.2016.041> (2016)
99. Trochim, W.M., Donnelly, J.P.: *Research Methods Knowledge Base*, vol. 2. Atomic Dog Publishing Macmillan Publishing Company, New York (2001)
100. Urbina, S.: *Fundamentos da testagem psicológica*. Artmed Editora (2009)
101. Usman, M., Britto, R., Borstler, J., et al.: Taxonomies in software engineering: a systematic mapping study and a revised taxonomy development method. *Inf. Softw. Technol.* **85**, 43–59 (2017). <https://doi.org/10.1016/j.infsof.2017.01.006>
102. Van Solingen, R., Berghout, E.W.: *The Goal/Question/Metric Method: A Practical Guide for Quality Improvement of Software Development*. McGraw-Hill, London (1999)
103. Wohlin, C.: Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*. Association for Computing Machinery, New York, NY, USA, EASE '14, pp. 1–10 (2014). <https://doi.org/10.1145/2601248.2601268>
104. Wohlin, C., Runeson, P., Höst, M., et al.: *Experimentation in Software Engineering*. Springer, New York (2012). <https://doi.org/10.1007/978-3-642-29044-2>
105. Yang, C.L., Yuan, C.W.T., Wang, T.Y., et al.: Knowing that you know what I know helps? Understanding the effects of knowledge transparency in online knowledge transfer. *Proc. ACM Hum. Comput. Interact.* **2**(CSCW), 1–21 (2018). <https://doi.org/10.1145/3274458>
106. Yin, R.K.: *Case Study Research: Design and Methods*, vol. 5. Sage, Thousand Oaks (2009)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.