The Awareness Assessment Model: measuring the awareness and collaboration support over the participant's perspective

Márcio J. Mantau[1], **Fabiane B.V. Benitti**[2]

[1]Departamento de Engenharia de Software
Universidade do Estado de Santa Catarina (UDESC) – Ibirama, SC – Brasil

[2]Departamento de Informática e Estatística
Universidade Federal de Santa Catarina (UFSC) – Florianópolis, SC – Brasil

marcio.mantau@udesc.br, fabiane.benitti@ufsc.br

Abstract. *[Context] Awareness has been a valuable concept in Collaborative Systems since its formation, being an essential part of groupware. The 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. [Problem] This is a multi-factorial problem, and finding a good starting point in the literature can be challenging for novice groupware designers; they must reinvent awareness from their own experience of what it is, how it works, and how it is used. [Goal] This work consists of establishing an assessment model for collaborative interfaces by analyzing the awareness mechanisms provided from the participant's viewpoint. Our awareness assessment model developed adopting the statistical technique Item Response Theory (IRT) and considers the participant's skill in understanding the awareness and the difficulty involved. [Results] The proposed assessment model allow us to measure the awareness support provided considering the collaboration, workspace, and contextual awareness perspectives. The results obtained were translated into an awareness support scale and three levels of quality were defined.*

1. Introduction

Needs such as connecting people, enabling individuals to collaborate, and supporting social interaction are part of the human essence, and collaborative systems are great environments to provide them. Applications such as email, instant messaging, chat forums, social networks, digital voicemail, and video conferencing applications are examples of daily tools that allow people to connect, interact and collaborate.

To provide the collaboration aspects, it must make available cues/information enabling participants to communicate, coordinate, and cooperate. This support involves a fundamental element of a collaborative system: the awareness [Dourish and Bellotti 1992]. Awareness has been an essential part of Collaborative Systems [Gross 2013, Tenenberg et al. 2016].

We consider awareness as the backbone of a collaborative environment and, through it, all collaborative concepts are archived. In this sense, we define awareness as a process that occurs at three basic levels of abstraction: representation, understanding, and projection. Providing an efficient awareness mechanism ensures a better understanding 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.

Over the last three decades, different awareness types have emerged in the literature. The works of [Seebach et al. 2011], [Antunes et al. 2014], [Gallardo et al. 2018], and [Mantau et al. 2017] present a broader list of awareness types. Detailed background about awareness origins, early ethnographic, and technology studies that brought up fundamental insights we found in [Gross 2013].

In recent efforts, [Mantau and Benitti 2022a], and [Mantau and Benitti 2022b] synthesized 10 years of literature on awareness support. First, we conducted a systematic mapping study and identified the main supporting awareness mechanisms. Second, adopting Bailey's taxonomy definition process [Bailey 1994], we refined these awareness mechanisms and developed a multi-dimensional taxonomy, represented in three main awareness dimensions, namely, collaboration, workspace, and contextual.

1.1. The Awareness Problem

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 [Niemantsverdriet et al. 2019]; because they must reinvent awareness from their own experience of what it is, how it works, and how it is used [Collazos et al. 2019].

Considering awareness at a high level, both people may differ in their understandings, and individual awareness may change as their background and received stimuli change. People have different capabilities in the representation, understanding, and projection of human actions through interface [Mantau et al. 2017]. Considering awareness as being, ultimately, an individual understanding, or even a mental state that an individual has about a certain object or environmental stimulus, we believe that the design, development, and in particular, evaluation approaches should, consequently, consider awareness from the participant's perspective.

1.2. Aims of Study

In this work, we evolve the findings presented by [Mantau and Benitti 2022b], providing contributions towards the development of an awareness assessment model that allows accessing the awareness, and consequently, the collaboration support through measuring awareness mechanisms from the participant's viewpoint. In this model, we consider the participant's skill in understanding the awareness information provided by the application and the difficulty involved in perceiving each awareness piece. In our assessment model, we assume the following assumptions:

i) Awareness is an individual understanding of a particular object or stimulus in the environment; It is the means available to interact with each other, and involves, from the participant's viewpoint, the representation (design mechanisms/elements that provide participants with cues about "what is going on") and the understanding/consciousness (state of being conscious of something);
ii) Collaboration is the result of the participant's understanding/consciousness; The consciousness allows individuals to project their actions;

iii) Awareness is intrinsically linked to the participant's own skills, whether in identifying, understanding, or projecting their actions; Different individuals may have different awareness, likewise, the participant's understanding differs over time.

2. The Awareness Assessment Model

The Awareness Assessment Model is developed specifically for evaluating collaborative systems which measure their quality by analyzing the awareness information provided by the application. Our assessment model is composed by:

i) 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 [Mantau and Benitti 2022b];
ii) The data collection and analysis tools present a set of support artifacts for conducting the collection and compilation of data obtained by interventions;
iii) The awareness assessment quality scales and measurement represent a set of useful elements to analyze the results obtained through assessment instruments and so, to classify the collaborative environment at a quality level through the participants' perspective.

The assessment is conducted by at least one examiner and by considering the participants' perspective as a data source, this instrument allows us to classify the collaborative environment into the awareness quality level.

2.1. Data Collection and Analysis Tools

The awareness assessment model contains a set of data collection instruments, through the application of ten awareness assessment questionnaires. In this approach, we considered the adoption of 75 specific awareness assessment items identified in the awareness taxonomy [Mantau and Benitti 2022b]. In order to reduce the number of assessment items for each participant, we used the balanced incomplete block design approach [Hinkelmann and Kempthorne 2005]. This approach is generally used when the number of treatments (assessment items, in our case) is hard to apply all at once.

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 λ. The BIB design must satisfy the following characteristics [Hinkelmann and Kempthorne 2005]:

i) Each block b have the same number of plots k (number of treatments), where $b.k = t.r$ and $b \geq t$;
ii) Every treatment is replicated r times in the design, where $r > k$;
iii) Each treatment occurs at most once in a block, and every pair of treatments occurs together λ times in the blocks, where $\lambda(t-1) = r(k-1)$; and
iv) 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 5 blocks of 15 assessment items each. Hence, we used questionnaires containing 2 blocks of items, totaling 30 questions. By applying the BIB method, we found a balanced incomplete block design composed of 10 distinct blocks (questionnaires). Table 1 presents the configuration of the treatments (t) and blocks of questionnaires items (b).

Table 1. Generated Balanced Incomplete Block Design

(a) Treatments (t)

(t)	**Awareness Assessment Items** (questions)
$t1$	$\{01, 09, 14, 17, 24, 30, 39, 42, 43, 46, 51, 56, 65, 67, 69\}$
$t2$	$\{02, 06, 08, 20, 28, 33, 34, 36, 38, 49, 53, 60, 64, 66, 68\}$
$t3$	$\{03, 12, 19, 21, 23, 27, 29, 31, 44, 47, 55, 57, 58, 61, 75\}$
$t4$	$\{04, 07, 11, 13, 15, 25, 45, 48, 50, 52, 54, 63, 70, 72, 73\}$
$t5$	$\{05, 10, 16, 18, 22, 26, 32, 35, 37, 40, 41, 59, 62, 71, 74\}$

(b) blocks (b)

block (b)		**block** (b)
$b1 = \{t1, t2\}$		$b6 = \{t3, t4\}$
$b2 = \{t3, t5\}$		$b7 = \{t2, t4\}$
$b3 = \{t1, t4\}$	...	$b8 = \{t4, t5\}$
$b4 = \{t1, t5\}$		$b9 = \{t1, t3\}$
$b5 = \{t2, t3\}$		$b10 = \{t2, t5\}$

The assessment model comprises 10 specific questionnaires that represent 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 GQM approach [Basili 1992], and guidelines presented by [Wohlin et al. 2012], [Wohlin 2014], and [DeVellis 2016]. The GQM paradigm defines a measurement model on three levels: conceptual level (goal); operational level (question); and quantitative level (metric and/or scale).

The questions were composed by combining the following structure: the component of the sentence (subject + predicate) present in Table 2 combined with the correspondent complement of each evaluation item (Table 3, 4, or 5, according to the awareness dimensions evaluated). In each applicable evaluation question, the participant selects an option according to how much you agree or disagree with each statement (4-points Likert scale). All assessment items were composed similarly. Tables 2 to 5 present the complete version of the questionnaire.

Table 2. Assessment Items' sentence components

Sentence	**Sentence component** (subject + predicate)
s1	"The collaborative environment allows me to identify"
s2	"When I'm interacting, I can identify"
s3	"By using the collaborative environment, I can identify"
s4	"When I am carrying out an activity together, I can identify"
s5	"During the interaction, I can identify"

3. Awareness Assessment Quality Scales

The awareness quality scales and awareness mechanisms measurement aims to classify the collaborative environment at a quality level through the participants' perspective. Our awareness assessment quality scales aim to access the Model's reliability and internal consistency, considering respectively, the Item Response Theory items' discrimination value. The awareness mechanisms measurement allows us to assess the general awareness quality of the collaborative environment, its presented design elements, goals, and awareness dimensions, through the estimate of the examinee's ability. In this sense, we assume the graded item response approach combined with the ability and item information functions proposed by [Samejima 1969] and [Baker and Kim 2004].

Table 3. Assessment items: workspace awareness dimension

Taxonomy [Mantau and Benitti 2022b]			**Questionnaire items** (object/complement)	
Workspace awareness	Involves activities	Goal	Q1	s1 + “the purpose of the tasks performed”
		Subject	Q2	s2 + “what are the artifacts or objects that are being changed”
		Content	Q3	s3 + “the current contents of shared resources”
		Motivation	Q4	s2 + “the motivation for the tasks performed”
		Time required	Q5	s3 + “the time needed/available to perform the task”
		Progress level	Q6	s2 + “the progress of tasks carried out together”
		Help needed	Q7	s4 + “how can I help the other participants”
		Evaluation	Q8	s5 + “the results obtained”
	Consider workflow	Authorship	Q9	s1 + “who is conducting the tasks/activities”
		Execution steps	Q10	s4 + “what steps/actions must be taken”
		Events and Actions	Q11	s5 + “what is happening in the environment”
		Change location	Q12	s3 + “the places where I can interact or perform the tasks”
		Related activities	Q13	s5 + “if there are other tasks that are related to the current scenario”
		Parallel activities	Q14	s5 + “if the other participants are engaged with the current task”
		Coordinated activities	Q15	s5 + “whether the task is being carried out in a coordinated manner”
		Adjusted activities	Q16	s5 + “how the current task is related to the current scenario (joint tasks)”
	Consider environment	Tools and Materials	Q17	s3 + “the tools and materials available to collaborate”
		Artifacts and Objects	Q18	s3 + “the presence of artifacts/objects needed to collaborate”
		Resources Availability	Q19	s3 + “the features available for collaborating”
		Critical elements	Q20	s3 + “if there are restrictions for carrying out the tasks”
		Virtual relationships	Q21	s3 + “the relationship between objects/environment resources”
	Provide understanding	Meaning	Q22	s4 + “ the meaning of the actions performed (what is happening)”
		Scenarios	Q23	s4 + “what are the next tasks that must be carried out”
		Sense-making	Q24	s4 + “the understanding of the other participants involved”
	Allow interaction	Feedback	Q25	s3 + “the response of actions performed by me”
		Feedthrough	Q26	s3 + “the response of the actions taken by the other participants”
		Backchannel feedback	Q27	s4 + “if the others are following the actions performed”
		Feedforward	Q28	s4 + “changes made by other participants”
	Consider relationship	Action control	Q29	s5 + “how other participants are controlling their actions/decisions”
		Access control	Q30	s3 + “who is controlling the environment, tasks, or shared resources”
		Access privileges	Q31	s3 + “the presence of access privileges in the shared environment”
		Control mechanisms	Q32	s1 + “if there are mechanisms to control access and how can I access them”

Table 4. Assessment items: collaboration awareness dimension

Taxonomy [Mantau and Benitti 2022b]			**Questionnaire items** (object/complement)	
Collaboration awareness	Allow identity	Identity	Q33	s3 + “the identity of the participants (who are they?)”
		Shared profile	Q34	s3 + “what information is being shared”
		Preferences	Q35	s1 + “the individual preferences of each participant ”
	Consider capabilities	Rules	Q36	s3 + “if there are different rules (and what they are) for each participant”
		Responsibilities	Q37	s5 + “the responsibilities that each participant can assume”
		Privileges	Q38	s5 + “what each participant can do, see or even control”
		Knowledge	Q39	s2 + “what I know about the current task/activity and how I can help”
		Influences	Q40	s2 + “what are the influences/decisions of each participant”
		Intentions	Q41	s4 + “my intentions and I can identify the others’ intentions”
	Provide Status	Availability	Q42	s3 + “the availability of each participant”
		Presence	Q43	s3 + “the presence of each participant in the environment”
		Activity level	Q44	s5 + “the level of activity/engagement of each participant”
		Status	Q45	s3 + “the current state/situation of each participant”
	Provide communication	Mode (sync or async)	Q46	s3 + “the working mode (synchronous or asynchronous)”
		Connectivity	Q47	s3 + “the network connectivity”
		Message delivery	Q48	s4 + “when messages are sent/received by other participants”
		Message delays	Q49	s5 + “if there are delays in sending/receiving messages”
		Interaction ways	Q50	s3 + “the means available to connect and interact with other participants”
		Turn-talking	Q51	s5 + “who is talking, exchanging ideas, or whose turn it is to speak”
		Conversation	Q52	s1 + “the means available to establish communication to other participants”
	Consider social	Expectations	Q53	s5 + “what are the expectations involving each participant”
		Emotional state	Q54	s5 + “the emotional state of each participant”
		Non-verbal cues	Q55	s5 + “the availability of non-verbal information for communication”

The quality scales have been developed adopting the statistical model denominated Item Response Theory (IRT), as presented by [Baker 2001]. The IRT refers to a family of mathematical models that relate observable variables (e.g., questionnaire items)

Table 5. Assessment items: contextual awareness dimension

Taxonomy [Mantau and Benitti 2022b]			**Questionnaire items** (object/complement)	
Contextual awareness	Consider spatiality	Location	Q56	s3 + "the physical/virtual location of other participants"
		Distance	Q57	s3 + "the distance of each participant in relation to the others"
		Restrictions	Q58	s3 + "whether there are space constraints involved (and what they are)"
		Places	Q59	s3 + "if there are different places for collaboration (and what they are)"
		Topology	Q60	s3 + "how the environment is configured"
		Attributes	Q61	s3 + "the attributes of the objects/resources or conditions of the environment"
		View	Q62	s1 + "what each participant can see"
		Reach	Q63	s1 + "the reach of each participant (where they can go, what they can access)"
		Orientation	Q64	s1 + "the orientation/direction of each participant"
		Movement	Q65	s4 + "the movement of each participant in the shared environment"
		Range of attention	Q66	s5 + "the level of attention needed to perform the task"
	Allow mobility	User modality	Q67	s1 + "if the system allows different access modes/devices (e.g. local/remote)"
		user mobility	Q68	s1 + "the user mobility (access by different devices)"
		Autonomy	Q69	s3 + "if there is a dependency between the application and the place of use"
	Provide navigation	Voice cues	Q70	s2 + "who is talking to whom (verbal communication)"
		Portholes/peepholes	Q71	s3 + "the means to peek the contents of tasks without accessing directly"
		Eye-gaze cues	Q72	s5 + "where each participant is looking"
		Map views	Q73	s1 + "the shared environment in a simplified form (thumbnail, map, or similar)"
		Viewports/Teleports	Q74	s3 + "the means to preview the tasks carried out by the other participants"
		Artifacts location	Q75	s3 + "where are the objects/artifacts or resources in the shared environment"

and hypothetical unobservable traits or aptitudes (e.g., awareness quality). They establish a link between the properties of items on an instrument, individuals responding to these items and the underlying trait being measured. Thus, there is a stimulus (item) that is presented to the subject, and he responds to it, and the response that the subject gives to the item depends on the level that the subject has in the latent trait or ability [Pasquali 2020]. IRT assumes that the latent construct (e.g. stress, knowledge, attitudes) and items of a measure are organized in an unobservable continuum.

At each ability level (θ) there will be a certain probability, denoted by P, that an examinee with that ability will give a correct answer to the item [Baker and Kim 2017]. In IRT, the function of ability $P(\theta)$, also represented by the item characteristic curve, 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 [Samejima 1969], where it is assumed that an item's response categories can be ordered with each other. On this model, the probability of a participant $j, \forall j \in J = \{1, 2, \ldots, m\}$ chose a score $k, \forall k \in K = \{0, 1, \ldots, n\}$, for a measurement item $i, \forall i \in I = \{1, 2, \ldots, o\}$ is given by the Equation 1.

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

Where $P_{i,k}$ is the probability that an individual j receives a score k in item i; e is a mathematical constant (the Euler's number, equals to $2.71828\ldots$); 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 $b_{i,1} \leq b_{i,2} \leq \cdots \leq b_{i,n}$.

Each item in a test will have its own item characteristic curve, and we considered two technical properties to describe it: the discrimination (a) and the difficulty (b). The discrimination parameter describes how well an item can discriminate (differentiate) the participants in relation to 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 in accordance with [Baker 2001]. A measurement instrument item is satisfactory in a measurement scale if the discrimination value $a \geq 0.65$, as presented in Table 6. Thus, measurement instrument items that have a discrimination parameter $a < 0.65$ are disregarded from the analysis, as they may not be correctly differentiating 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.

Table 6. Item discrimination parameter values

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

To position the items on the scale and identify the categories of the scale (quality levels), we considered the probability parameter $P_{i,k}(\theta) \leq 0.5$ and scale $(0, 1)$ [de Andrade et al. 2000]. This scale is widely used by IRT to represent, respectively, the mean value and the standard deviation of the individual abilities of the population, and in this case, the values of parameter b vary between -2 and $+2$. Regarding the parameter a, values between 0 and $+2$ are expected, and the most appropriate values would be those greater than 0.65.

3.1. Awareness Mechanisms Measurement

On the IRT, the evaluation information is defined in terms of item information functions $I_i(\theta)$, which is a measure of how good responses in that category estimate the examinee's ability [Baker and Kim 2004]. In our model, we assume the graded item response approach, where each item has been divided into n ordered response categories.

For each awareness dimension $d, \forall d \in D = \{workspace, collaboration, contextual\}$ and considering the applicable awareness goals g, $\forall g \in G_d = \{\forall g \mid g\ is\ a\ goal \in awareness\ dimension\ d\}$, their related measurement items i, $\forall i \in I_{gd} = \{\forall i \mid i\ is\ an\ measurement\ item \in awareness\ dimension\ d\ and\ goal\ g\}$, and item scores k, denoting an arbitrary category $\forall k \in K = \{0, 1, \ldots, n\}$, where n is the number of response categories for item i, we calculate The item's information $I_i(\theta)$, for each applicable questionnaire item i, considering the awareness goal g of the awareness dimension d, by using the item information function proposed by [Samejima 1969] (Equation 2).

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

It is essential to highlight that Equation 2 calculates 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.

4. Model Validation

The validation process was performed in two main steps. First, during the assessment model construction, we expose the model's artifacts to experts using the expert panel approach [Beecham et al. 2005]. Second, the reliability and validity of the measurement instrument were identified based on data collected through two case studies conducted.

4.1. Expert Panel Validation

To improve the assessment model, we expose the model's artifacts to expert appreciation through the expert panel approach [Beecham et al. 2005]. 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 [Nickerson et al. 2013]. 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 both to address whether a purposeful and unambiguous determination is possible, by evaluating the practical applicability and to demonstrate whether a clear definition of its elements can be made [Strasser 2017]. This approach also allows reflecting on the current state of research on an object [Khalilijafarabad et al. 2016], to discover similarities and differences between studies on this type of object [Agogo and Hess 2018], and to identify potential research gaps [Hummel et al. 2016].

Based on the Goal Question Metric approach [Basili 1992, Van Solingen and Berghout 1999], 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 Table 7.

In this step, we are exposing our assessment model to expert evaluation, like awareness, collaborative systems, and HCI researchers, in order to identify its suitability for evaluating collaborative environments. After this refinement, we reviewed the exposed artifacts and started the large-scale model evaluation process, by planning and executing two case studies, detailed in Section 4.2.

4.1.1. Expert Panel Results

From May to June 22, we presented the data collection artifacts (questionnaire) to expert opinion and five expert assessments of the initial model were obtained. Figure 1 summarizes the obtained results.

Overall, the evaluation model received a good rating from the expert's perspective. On a gradual scale, from 1 (strongly disagree) to 5 (strongly agree), the assessment

Table 7. Expert panel questionnaire items

Validation aspect		Questionnaire items	Response type
Demographic	D1	Indicate your expertise in relation to the following: a) Awareness; b) Collaborative systems; c) Human-Computer Interaction;	gradual scale (from novice to expert)
Purposeful	M1	Relevance: all statements in the representation are relevant to the problem	gradual scale (from strongly disagree to strongly agree)
Correctness	M2	Correctness: all statements in the representation are correct	
	M3	Understandability: the purpose, concepts, and structure of the reference model are clear to the users	
Applicability	M4	Authenticity: the representation gives a true account of the domain	
	M5	Generality: the reference model is usable in different cases	
	M6	Usability: users can easily apply the reference model	
	M7	Completeness: the representation contains all statements about the domain that are correct and relevant	
General evaluation	M8	Purposeful: write your impressions, strengths, weaknesses, suggestions, or comments relevant to the purposeful aspect	Plain text
	M9	Unambiguous: write your impressions, strengths, weaknesses, suggestions, or comments relevant to the unambiguous determination aspect	
	M10	Applicability: write your impressions, strengths, weaknesses, suggestions, or comments relevant to the applicability aspect	

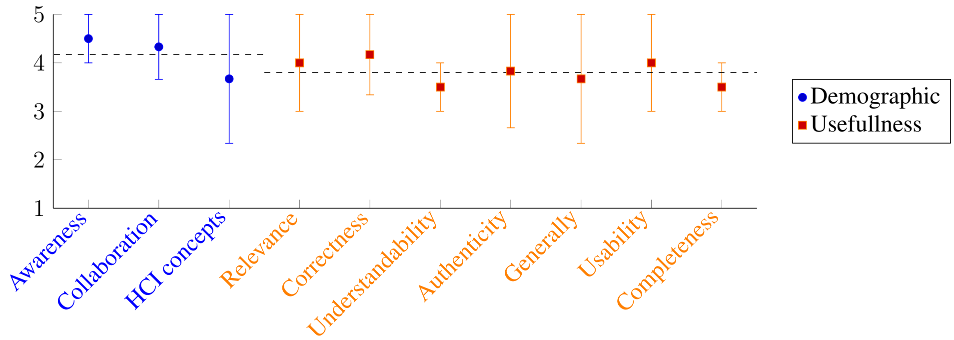


Figure 1. Expert panel questionnaire results

items M1 to M7 received values over 3,5 (average 3,8). Although the small sample of specialists, all reported having a good experience regarding awareness, collaboration, and HCI concepts, corroborating with 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, the received feedback demonstrates a concern regarding the clarity of the specification and whether the model contains all statements about the domain or if they can be applied to the same environment. We thought that, depending on the domain of the collaborative system, not all aspects will be applied – and this will not necessarily be a weak point of the model. For example, if a system focuses on performing synchronous or asynchronous work, the awareness information may differ. In some points, the awareness mechanisms require balancing the need of presenting proper awareness support while dealing with information overload or intrusiveness, for example.

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 points scale, does not corroborate with our intended analysis model.

4.2. Case Studies

After the expert panel refinement, we reviewed the exposed artifacts and started the model evaluation process, by planning and executing a set of case studies. In this section, we present the preliminary results obtained through the realization of two case studies.

In the first case study, we evaluated virtual collaboration environments, intended for communication/interaction between two or more people simultaneously. For example, conference environments, videoconferencing, virtual events, webinars, etc. In these environments, in order to have a satisfactory interaction, it is necessary to provide awareness cues such as the participants' profile, capabilities, status, forms of communication, and social aspects.

In the second case study, we evaluate the collaboration aspects provided by the Moodle platform. Moodle is a collaborative learning platform designed to create learning environments, aimed at educators and students. For collaborative learning to occur satisfactorily in this environment, the application needs to provide awareness cues, like, the participants' profile, communication and interaction resources, and spaces to share artifacts, objects, and materials.

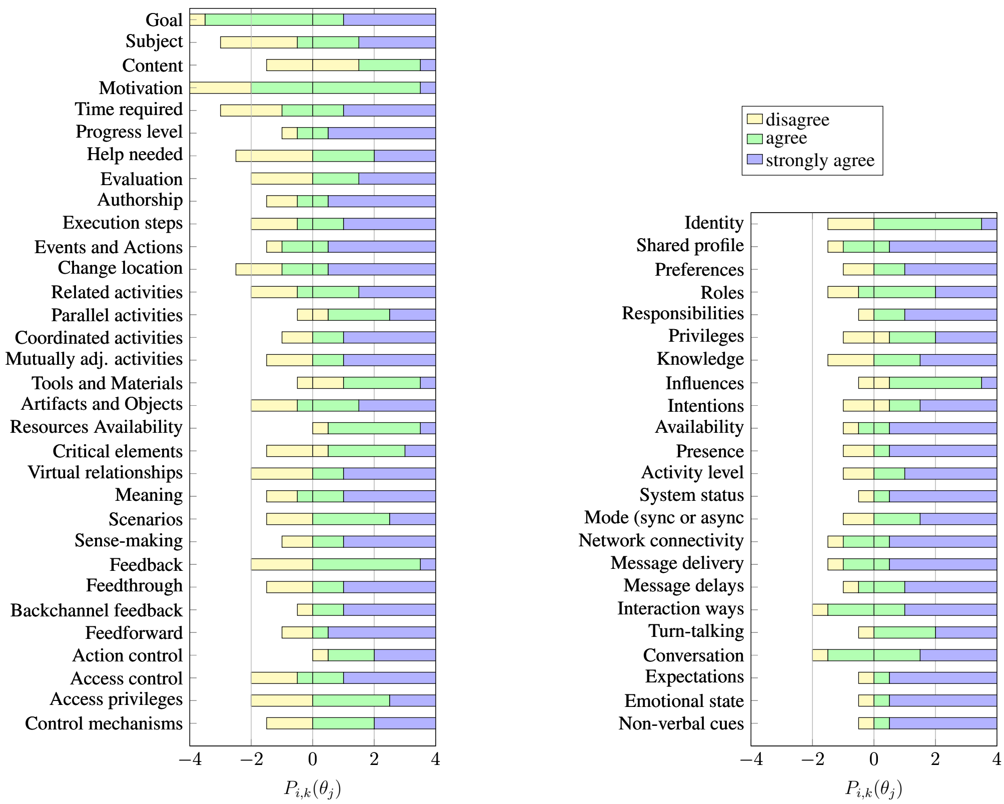
4.2.1. Case studies results

For both case studies, we compiled the blocks (treatments) into 10 different test books and then set up an online questionnaire (Google Forms) to collect participant feedback. We obtained the voluntary participation of 71 students in the first and 54 in the second case study, totaling 125 observations. Due to the small number of observations obtained, we combined in our analysis the observations of both scenarios to generate the TRI scale.

Figures 2a, 2b, and 3a show the probability scales generated through the IRT model. For each awareness mechanism described in our taxonomy [Mantau and Benitti 2022b], we calculated the relationship between the probability of each response item (from strongly disagree to strongly agree) in relation to the individual's ability scale. In this representation, the probability of the individual evaluating each of the items considers the difficulty/skill that the participant has, that is, elements that are more difficult to be understood require a greater skill scale for their assessment.

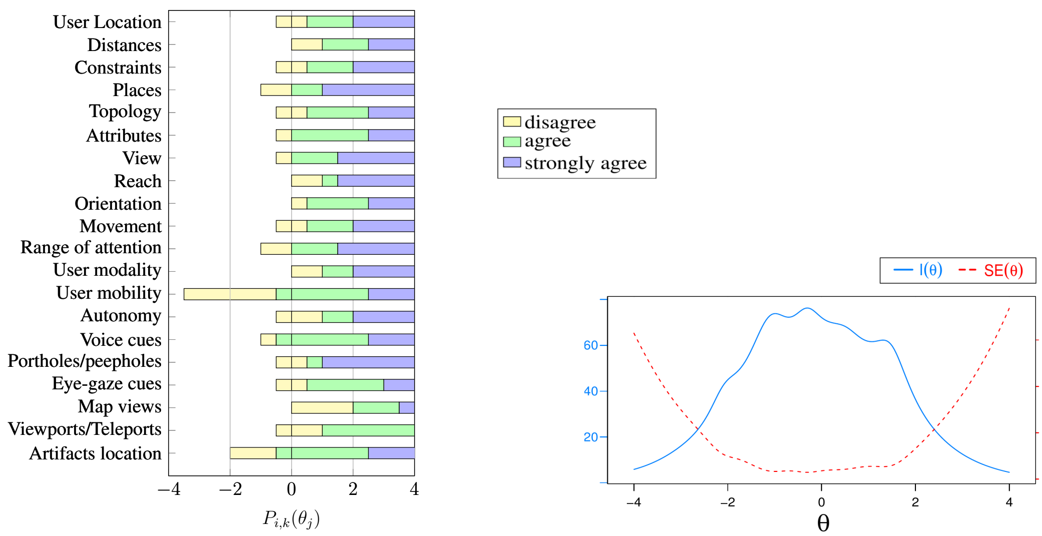
The total information curve of the awareness mechanisms' support, Figure 3b, shows the sum of information for all assessment items. The adequate representation of this scale evidences the reliability of the instrument. This graph represents the region of the ability scale (θ_j) where the participant (j) is able to 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 are able to identify the mechanisms well ($\theta_j > 1$).

The total instrument information curve also shows the standard error (SE) behavior relevant to instrument accuracy. It is observed that the SE curve (red dotted line) reaches its minimum value exactly 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 region of the scale where the information curve is greater than the standard error curve, that is, between the scale levels [-2.5, +2.5].



(a) Workspace awareness **(b) Collaboration awareness**

Figure 2. Ability level scales



(a) Contextual awareness **(b) Test information and standard errors**

Figure 3. Ability level scales and Test information

Based on the positioning of items throughout the scale, three levels of quality are defined: low quality, good quality, and excellent quality. Table 8 exemplifies the collaborative environment classification through our awareness quality scales. The IRT calculates a participant's score and positions it on the defined ability scale. However, we are interested in classifying the collaborative environment, thus, we calculate here the average of the provided scores of all participants involved.

The positioning of respondents over the awareness scale is obtained using the theta parameter (θ), which represents the competence score of each subject. In both case studies, the collaborative environments provided good awareness support: $\theta = 0.45$ in the videoconferencing and $\theta = -0.40$ in the Moodle scenario.

Table 8. Awareness mechanisms quality scales

	Quality level	Level description
Workspace awareness	Low ($\theta < -1$)	The collaborative environment rarely provides workspace capabilities and hardly provides 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 ($-1 \leq \theta \leq 1$)	The collaborative environment sometimes supports workspace capabilities and presents some information about the activities, environment, and workflow. The interaction and understanding of artifacts and objects shared are possible, although, usually do not present a fully attractive design or good operability.
	Excellent ($\theta > 1$)	At this level, the collaborative environment provides full support for workspace capabilities and provides information about the activities, environment, and workflow. The environment provides efficient interaction and understanding of artifacts and objects shared in the workspace. In terms of usability, the workspace elements present excellent operability.
Collaboration awareness	Low ($\theta < -1$)	The collaborative environment rarely provides social interaction and collaboration aspects. The environment hardly provides status, and identity information, nor does it consider the participant's capabilities. Due to these limitations, the collaboration aspects are limited.
	Good ($-1 \leq \theta \leq 1$)	The collaborative environment sometimes presents social interaction and collaboration aspects. The environment provides moderate status, and identity information. Sometimes is considered the participant's capabilities. Often the awareness information is considered relevant to the participant's interests and, usually, they recognize that the content helps in the collaboration process.
	Excellent ($\theta > 1$)	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 that is, it has clear rules and is easy to interact with others.
Contextual awareness	Low ($\theta < -1$)	The collaborative environment hardly considers the contextual perspective or the group members' mobility. Environmental navigation or spatiality are rarely allowed. Due to these limitations, contextual interaction is limited.
	Good ($-1 \leq \theta \leq 1$)	The collaborative environment provides moderated access to contextual information. Environmental navigation and spatiality aspects are partially reached by participants. The environment provides some operability over participants' contextual information.
	Excellent ($\theta > 1$)	At this level, the collaborative environment provides clear and easy access to contextual information. Environmental navigation and spatiality aspects are fully reached by participants. The environment presents excellent operability over participants' contextual information.

5. Conclusion

Awareness is an individual understanding, a mental state, about a certain object or environmental stimulus, and involves, from the participant's viewpoint, the representation and understanding/consciousness process. Furthermore, the awareness process depends on the participant's skills, whether in identifying, understanding, or projecting their actions; different individuals may have different awareness, likewise, the participant's understanding differs over time. Our awareness assessment model uses the IRT model, which fits the awareness and, consequently, collaboration support into an individual perspective.

The instrument's content validity, that is, the investigation that the instrument's items are correctly measuring essential aspects of the phenomenon, was observed together with the specialists of the construct. The measurement instrument presented evidence of validity and reliability, verified by the IRT model. Validity evidence was verified in the elaboration stage and by adjusting the items to the Graded Response Model [Samejima 1969]. Reliability evidence was verified using the Test Information and Standard Error curves.

References

Agogo, D. and Hess, T. J. (2018). "how does tech make you feel?" a review and examination of negative affective responses to technology use. *European Journal of Information Systems*, 27(5):570–599.

Antunes, P., Herskovic, V., Ochoa, S. F., and Pino, J. A. (2014). Reviewing the quality of awareness support in collaborative applications. *Journal of Systems and Software*, 89(1):146–169.

Bailey, K. D. (1994). *Typologies and taxonomies: An introduction to classification techniques*, volume 102. Sage.

Baker, F. B. (2001). *The basics of item response theory*. ERIC.

Baker, F. B. and Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques*. CRC press.

Baker, F. B. and Kim, S.-H. (2017). *The basics of item response theory using R*. Springer.

Basili, V. (1992). Software modeling and measurement: The goal/question/metric paradigm. *Technical Report*.

Beecham, S., Hall, T., Britton, C., Cottee, M., and Rainer, A. (2005). Using an expert panel to validate a requirements process improvement model. *Journal of Systems and Software*, 76(3):251–275.

Collazos, C. A., Gutierrez, F. L., Gallardo, J., Ortega, M., Fardoun, H. M., and Molina, A. I. (2019). Descriptive theory of awareness for groupware development. *Journal of Ambient Intelligence and Humanized Computing*, 10(12):4789–4818.

de Andrade, D. F., Tavares, H. R., and da Cunha Valle, R. (2000). Teoria da resposta ao item: conceitos e aplicações. *ABE, Sao Paulo*.

DeVellis, R. F. (2016). *Scale development: Theory and applications*, volume 26 of *Applied Social Research Methods 26*. Sage Publications, Inc, 4th edition.

Dourish, P. and Bellotti, V. (1992). Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work*, CSCW '92, pages 107–114, New York, NY, USA. Association for Computing Machinery.

Gallardo, J., Bravo, C., and Molina, A. I. (2018). A framework for the descriptive specification of awareness support in multimodal user interfaces for collaborative activities. *Journal on Multimodal User Interfaces*, 12(2):145–159.

Gross, T. (2013). Supporting effortless coordination: 25 years of awareness research. *Computer Supported Cooperative Work*, 22(4-6):425–474.

Hinkelmann, K. and Kempthorne, O. (2005). *Design and analysis of experiments: Advanced experimental design*. Wiley-Interscience.

Hummel, D., Schacht, S., and Maedche, A. (2016). Determinants of multi-channel behavior: Exploring avenues for future research in the services industry. In *ICIS*.

Khalilijafarabad, A., Helfert, M., and Ge, M. (2016). Developing a data quality research taxonomy-an organizational perspective. In *ICIQ*, pages 176–186.

Mantau, M. J. and Benitti, F. B. V. (2022a). Awareness support in collaborative system: Reviewing last 10 years of cscw research. In *5th IEEE International Conference on Computer Supported Cooperative Work in Design, CSCWD 2022, Hangzhou, China, May 4-6, 2022*, pages 564–569. IEEE.

Mantau, M. J. and Benitti, F. B. V. (2022b). Towards an awarenes taxonomy. In *25th IEEE International Conference on Computer Supported Cooperative Work in Design, CSCWD 2022, Hangzhou, China, May 4-6, 2022*, pages 495–500. IEEE.

Mantau, M. J., Berkenbrock, C. D. M., and Berkenbrock, G. R. (2017). Análise de requisitos de percepção em um groupware móvel síncrono. *Revista de Sistemas de Informação da FSMA*, 1(19):16–33.

Nickerson, R. C., Varshney, U., and Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, 22(3):336–359.

Niemantsverdriet, K., Essen, H. V., Pakanen, M., and Eggen, B. (2019). Designing for awareness in interactions with shared systems: The dass framework. *ACM Transaction on Computer-Human Interaction*, 26(6):1–41.

Pasquali, L. (2020). *TRI–Teoria de resposta ao item: Teoria, procedimentos e aplicacoes*. Editora Appris.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*.

Seebach, C., Beck, R., and Pahlke, I. (2011). Situation awareness through social collaboration platforms in distributed work environments. In *International Conference on Information Systems 2011, ICIS 2011*, volume 1, pages 1–22, Shanghai, China.

Strasser, A. (2017). Delphi method variants in information systems research: Taxonomy development and application. *Electronic Journal of Business Research Methods*, 15(2):pp120–133.

Tenenberg, J., Roth, W.-M., and Socha, D. (2016). From i-awareness to we-awareness in cscw. *Computer Supported Cooperative Work (CSCW)*, 25(4):235–278.

Van Solingen, R. and Berghout, E. W. (1999). *The Goal/Question/Metric Method: a practical guide for quality improvement of software development*. McGraw-Hill.

Wohlin, C. (2014). 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*, EASE '14, pages 1–10, New York, NY, USA. Association for Computing Machinery.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). *Experimentation in software engineering*. Springer Science & Business Media.