UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS REITOR JOÃO DAVID FERREIRA LIMA
GRADUATE PROGRAM IN COMPUTER SCIENCE

Márcio José Mantau

AN AWARENESS ASSESSMENT MODEL FOR COLLABORATIVE INTERFACES

Florianópolis, Santa Catarina – Brazil
2024

Márcio José Mantau

AN AWARENESS ASSESSMENT MODEL FOR COLLABORATIVE INTERFACES

Thesis submitted to the Graduate Program in Computer Science of Universidade Federal de Santa Catarina for degree acquirement in Doctor of Science degree in Computer Science.
Supervisor: Fabiane Barreto Vavassori Benitti, PhD.

Florianópolis, Santa Catarina – Brazil
2024

Legal Notes:

There is no warranty for any part of the documented software. The authors have taken care in preparing this thesis but make no expressed or implied warranties of any kind and assume no responsibility for errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising from using the information or programs contained here.

Cataloging at source by the University Library of the Federal University of Santa Catarina.
File compiled at 17:37h of the day Wednesday 5th June, 2024.

Márcio José Mantau

An Awareness Assessment Model for Collaborative Interfaces / Márcio José Mantau; Supervisor, Fabiane Barreto Vavassori Benitti, PhD. – Florianópolis, Santa Catarina – Brazil, April 9th, 2024.

308 p.

Thesis – Universidade Federal de Santa Catarina, INE – Department of Informatics and Statistics, CTC – Technological Center, Graduate Program in Computer Science.

Includes references

1. Awareness, 2. Collaborative systems, 3. Assessment model, I. Fabiane Barreto Vavassori Benitti, PhD. II. Graduate Program in Computer Science III. An Awareness Assessment Model for Collaborative Interfaces

CDU 02:141:005.7

Márcio José Mantau

AN AWARENESS ASSESSMENT MODEL FOR COLLABORATIVE INTERFACES

This work at PhD level was evaluated and approved on April 9, 2024, by the examination board composed of the following members:

Prof. Fabiane Barreto Vavassori Benitti, PhD.
Universidade Federal de Santa Catarina – UFSC

Prof. Jean Carlo Rossa Hauck, PhD.
Universidade Federal de Santa Catarina – UFSC

Prof. Carla Diacui Medeiros Berkenbrock, PhD.
Universidade do Estado de Santa Catarina – UDESC

Prof. Jose Maria Nazar David, PhD.
Universidade Federal de Juiz de Fora – UFJF

We certify that this is the original and final version of the conclusion work that was judged suitable for obtaining the title of Doctor in Computer Science.

ICP-Edu
Documento assinado digitalmente
Marcio Bastos Castro
Data: 06/06/2024 08:03:20-0300
CPF: ***.707.720-**
Verifique as assinaturas em https://v.ufsc.br

Márcio Bastos Castro, PhD.
Coordinator
Graduate Program in Computer Science

GOV.BR
Documento assinado digitalmente
FABIANE BARRETO VAVASSORI BENITTI
Data: 05/06/2024 16:40:00-0300
CPF: ***.986.440-**
Verifique as assinaturas em https://v.ufsc.br

Fabiane Barreto Vavassori Benitti, PhD.
Supervisor

Florianópolis, Santa Catarina – Brazil, April 9th, 2024.

This work is dedicated to my wife Amanda Cristina,
my daughters Alice Cristina and Julia Manuela,
and my son José Eduardo (in memoriam).

ACKNOWLEDGEMENTS

LORD my God, I take refuge in You; You have always given me the strength and courage to overcome adversities.

I thank my wife, Amanda Cristina, for understanding my absences and anxieties. I am grateful for her affection and support during the time I spent writing this thesis. You did not let me weaken, but you continually encouraged me.

I want to thank my daughters, Alice Cristina and Julia Manuela, for their sweetness and affection; I hope we will have more moments together at the end of this project. My dear son José Eduardo *(in memoriam)*, you will always have a special place in my mind, heart, and soul.

To my parents, who are an example of life, struggle, determination, honesty, and work, who have encouraged me to study since I was a child.

I want to thank my supervisor for her example as a person and professional, for her extreme competence, patience, confidence, questions, encouragement, and insight.

I want to thank the PPGCC professors for their commitment and dedication.

To others who contributed in some way to this work, we thank you very much!

The LORD is my shepherd, I shall not be in want.
– Psalm 23.1.

RESUMO

[Contexto] Nós definimos percepção *(awareness)* como um conjunto de processos nos quais reconhecemos, organizamos e damos sentido aos estímulos que recebemos do nosso ambiente. Este conceito tem sido valioso em sistemas colaborativos desde a sua formação, e pesquisas relacionadas acompanharam a evolução de todo o campo nas últimas décadas. Consideramos a percepção como a espinha dorsal de um ambiente colaborativo, na qual toda a colaboração é alcançada. [Problema] Segundo a literatura, a percepção é um conceito bem conhecido, mas ainda não totalmente alcançado. A percepção continua a ser difícil de compreender e pesquisas devem alcançar uma melhor compreensão do apoio concebendo e testando novas tecnologias que ofereçam um suporte adequado. Neste ponto, duas questões permanecem em aberto: primeiro, como entender a percepção e como ela deve ser encarada em aplicações colaborativas e, segundo, como estabelecer uma base para avaliar interfaces colaborativas focadas na percepção. [Objetivo] Este trabalho consiste em estabelecer um modelo de avaliação de percepção para interfaces colaborativas a partir da perspectiva dos participantes, analisando informações de percepção fornecidas pela aplicação. [Método] Esta pesquisa foi realizada em quatro etapas. Primeiro, realizamos um estudo de mapeamento sistemático para identificar as estratégias mais avançadas adotadas no desenvolvimento e avaliação da percepção e como elas são usadas no contexto colaborativo. Em segundo lugar, elaboramos uma taxonomia que contempla aspectos de percepção e colaboração necessários ao trabalho cooperativo. Terceiro, com base nos fundamentos da literatura, elaboramos um modelo de avaliação. Por fim, nós validamos o modelo de avaliação sob duas perspectivas: *i)* para melhorar o modelo proposto, expomos os artefatos do modelo à apreciação de seis especialistas por meio da abordagem de painel de especialistas. *ii)* para verificar a confiabilidade e validade de construto do instrumento, nós executamos três estudos de caso por meio de uma avaliação em larga escala do modelo proposto. Ao todo, foram coletadas 820 observações de usuários e envolveu a participação de 25 examinadores. [Resultado] Realizamos um estudo de mapeamento sistemático para identificar o apoio à percepção no contexto do sistema colaborativo e compilamos os resultados em uma taxonomia multidimensional de três dimensões principais de percepção: colaboração, espaço de trabalho e contextual, que abrange 75 mecanismos de percepção descritos. Com base nestes resultados, apresentamos um novo método de avaliação para percepção e apoio à colaboração centrado na perspectiva do participante, desenvolvendo um instrumento de medição baseado na Teoria de Resposta ao Item – TRI. [Conclusão] A metodologia permitiu-nos construir e interpretar uma escala de qualidade para avaliar o suporte à percepção na perspectiva do participante. Os cenários avaliados corroboram a validade do modelo. Primeiro, os resultados do painel de especialistas sugerem que o modelo possui validade de conteúdo. Segundo, os cenários de estudo de caso demonstraram: *i)* indicadores adequados do ponto de vista dos dados demográficos e da parametrização da TRI; *ii)* a aplicabilidade da escala de consciência na perspectiva do participante; e *iii)* a replicabilidade do modelo para diferentes cenários e contextos.

Palavras-chaves: Percepção; Sistemas Colaborativos; Modelo de avaliação.

RESUMO EXPANDIDO

INTRODUÇÃO

Necessidades como conectar pessoas, permitir que os indivíduos colaborem e apoiar a interação social fazem parte da essência humana, e os sistemas colaborativos são ótimos ambientes para atendê-las. Um sistema colaborativo, também conhecido como software colaborativo ou *groupware* (termo em inglês), é um sistema baseado em computador que suporta duas ou mais pessoas envolvidas em uma tarefa ou objetivo comum e fornece uma interface para um ambiente compartilhado (ELLIS; GIBBS; REIN, 1991). Esses sistemas ajudam os membros de um grupo a trabalhar juntos e permitem que os membros do grupo compartilhem informações e usem essas informações para apoiar o trabalho conjunto (GEORGE, 2003). Em essência, sistemas colaborativos permitem que os membros do grupo se comuniquem, coordenem o seu trabalho e cooperem. A colaboração ocorre quando duas ou mais pessoas, entidades ou organizações trabalham juntas para concluir uma tarefa ou atingir um objetivo.

Para fornecer suporte adequado à colaboração, o ambiente deve disponibilizar mecanismos (elementos, pistas, informações) que possibilitem que participantes se comuniquem, coordenem e cooperem. Este apoio envolve um elemento fundamental: a percepção (do inglês, *awareness*) (DOURISH; BELLOTTI, 1992). A percepção pode ser definida como um conjunto de processos nos quais nós reconhecemos, organizamos e damos sentido aos estímulos recebidos do ambiente. Este conceito tem sido valioso em sistemas colaborativos desde a sua formação (TENENBERG; ROTH; SOCHA, 2016), e pesquisas relacionadas acompanharam a evolução de todo o campo nas últimas décadas (GROSS, 2013). Ainda, a percepção pode ser considerada a espinha dorsal de um ambiente colaborativo, na qual toda a colaboração é alcançada. Assim, mecanismos de percepção eficientes apoiam uma melhor compreensão e projeção das ações futuras; em contrapartida, sua ausência prejudica a compreensão e impossibilita que os participantes projetem o seu trabalho.

Do ponto de vista do suporte à percepção, este trabalho assume que:

i) A percepção deve ser interpretada como sendo parte da compreensão individual de um determinado objeto ou estímulo ambiental e, do ponto de vista do participante, o meio disponível para interagir envolve a representação (mecanismos ou elementos que fornecem aos participantes pistas sobre "o que está acontecendo") e a compreensão (ou consciência) de algo;

ii) A percepção está intrinsecamente ligada às habilidades do participante em identificar, compreender ou projetar suas ações; assim sendo, sob o ponto de vista dos mecanismos de colaboração, indivíduos podem ter percepções diferentes e, da mesma forma, a compreensão do participante pode variar ao longo do tempo;

iii) A colaboração é resultante da compreensão/consciência do participante; logo, a consciência é fator determinante, que permite aos indivíduos projetarem suas ações.

Segundo a literatura, a percepção é um conceito bem conhecido, mas ainda não totalmente alcançado (GROSS, 2013); pesquisas voltadas à uma melhor compreensão dos conceitos envolvidos e seu adequado suporte em ambientes colaborativos fazem-se necessárias. Neste ponto, duas questões permanecem em aberto: primeiro, como entender a percepção e como ela deve ser encarada em aplicações colaborativas e, segundo, como estabelecer uma base para avaliar interfaces colaborativas sob a perspectiva do suporte à percepção.

Em relação ao entendimento do que é a percepção, pode-se considerá-la um problema multifatorial, pois, em um nível mais elevado [ou abstrato], as pessoas podem diferir na sua compreensão, e a percepção individual pode mudar à medida que mudam os seus antecedentes e os estímulos recebidos. De fato, pessoas possuem diferentes habilidades para representar, compreender e projetar ações humanas por meio de interfaces; de mesmo modo, fatores sociotécnicos como motivação, conhecimento e objetivos dos participantes influenciam a interação.

Segundo, do ponto da avaliação da percepção, identifica-se na literatura diferentes estratégias desenvolvidas para auxiliar projetistas (*designers*) a implementar mecanismos de percepção nas etapas de projeto e desenvolvimento de aplicações colaborativas (SANTOS; FERREIRA; PRATES, 2012; STEINMACHER; CHAVES; GEROSA, 2013; LOPEZ; GUERRERO, 2017; GALLARDO; BRAVO; MOLINA, 2018; COLLAZOS *et al.*, 2019; BRAVO *et al.*, 2023). Por outro lado, as abordagens mais comuns são projetadas para um contexto específico e não se concentram na avaliação desses mecanismos nem no suporte fornecido sob o ponto de vista do usuário. Embora se encontre vários métodos para avaliar sistemas colaborativos, sejam eles pré-existentes, novos, *ad-hoc* ou adaptados, poucos estudos apresentam métodos ou processos que forneçam uma avaliação com foco no suporte à percepção.

OBJETIVOS

Este trabalho consiste em estabelecer um modelo de avaliação de percepção para interfaces colaborativas a partir da perspectiva dos participantes, analisando informações de percepção fornecidas pela aplicação. Como objetivos específicos,destaca-se:

i) Identificar o estado da arte das abordagens (modelos, metodologias ou processos) adotadas no projeto, desenvolvimento e, principalmente, avaliação de sistemas colaborativos, abordando conceitos de percepção e colaboração;

ii) Identificar os elementos de percepção necessários e que as interfaces colaborativas devem suportar e como podemos representá-los conceitualmente;

iii) Estabelecer uma taxonomia de percepção que permita que aplicações colaborativas alcancem, por meio do fornecimento adequado destes mecanismos, aspectos de colaboração necessários para o trabalho colaborativo;

iv) Elaborar um modelo de avaliação de percepção para avaliar aplicações colaborativas a partir da perspectiva dos participantes por meio do acesso ao suporte de percepção fornecido;

v) Estabelecer uma escala de percepção global intercambiável para diferentes cenários;

vi) Elaborar um conjunto de artefatos de avaliação que orientem todo o processo avaliativo, desde instrumentos de coleta de dados, ferramentas de análise e construção e interpretação de escalas de percepção;

vii) Validar o modelo de avaliação de percepção em diferentes cenários.

METODOLOGIA

O método de pesquisa adotado foi inspirado pela abordagem conhecida como *Design Sceince Research* (DSR) (BICHLER, 2006). Esta pesquisa foi realizada em quatro etapas.

Primeiro, realizou-se um estudo de mapeamento sistemático (PETTICREW; ROBERTS, 2006; KITCHENHAM; CHARTERS, 2007) para identificar as estratégias adotadas no desenvolvimento e avaliação da percepção e como elas são usadas no contexto colaborativo. A análise do estado da arte visa identificar as abordagens existentes adotadas no desenvolvimento e avaliação de sistemas colaborativos, abordar conceitos de awareness e colaboração (modelo 3C) e identificar desafios e limitações relacionados. O processo de mapeamento sistemático é dividido em três fases: definição, execução e análise.

Na fase de definição, são identificados os objetivos da pesquisa e definido o protocolo de mapeamento sistemático. O protocolo especifica as questões de pesquisa e os procedimentos utilizados para conduzir a revisão, como definição das fontes de dados, string de busca, inclusão/exclusão e critérios de qualidade. A fase de execução consiste na busca, identificação e seleção de estudos relevantes segundo os requisitos definidos no protocolo. Durante a fase de análise, extraiu-se os dados utilizando o formulário de extração de dados.

Em segundo lugar, foi elaborado uma taxonomia que contempla aspectos de percepção e colaboração necessários ao trabalho cooperativo. O método de definição da taxonomia utilizado consiste em quatro fases: planejamento, identificação, projeto e construção, e validação (USMAN *et al.*, 2017; SZOPINSKI; SCHOORMANN; KUNDISCH, 2019).

Na fase de planejamento são definidos o contexto da taxonomia e sua configuração inicial, abrangendo a definição das meta características e as condições finais objetivas e subjetivas. Na fase de identificação, os dados para definir a nova taxonomia foram coletados a partir dos resultados do mapeamento sistemático. Nesta etapa, os termos foram coletados e as redundâncias e inconsistências foram identificadas e removidas por meio de um processo de controle terminológico. Ainda, utilizando uma análise fenética (NICKERSON; VARSHNEY; MUNTERMANN, 2013), os elementos foram classificados por similaridade. No final da fase de projeto e construção, verificou-se se todas as condições finais objetivas e subjetivas foram atendidas. Por fim, na etapa de validação, para garantir a utilidade da taxonomia e reforçar sua confiabilidade, cenários ilustrativos e estudos de caso foram utilizados.

Terceiro, com base nos fundamentos da literatura, elaborou-se um modelo de avaliação. Este modelo de avaliação é desenvolvido explicitamente para avaliar sistemas colaborativos, que tem sua qualidade medida por meio da análise das informações de percepção fornecidas pela aplicação. Considerando a percepção dos participantes como fonte de dados, este modelo de avaliação permite classificar o ambiente colaborativo em uma escala de qualidade. O modelo desenvolvido compreende o processo de Avaliação do suporte à percepção e uma visão conceitual.

O processo de avaliação da percepção é baseado em um conjunto de diretrizes de HCI e é inspirado no processo de avaliação definido pela norma ISO/IEC 25040:2011. Este consiste em três fases principais: planejamento, execução e reflexão. Esse processo é realizado pelo pesquisador/examinador, que avalia as interfaces colaborativas analisando as informações [ou mecanismos] de percepção fornecida pela aplicação. Este processo envolve a participação de uma amostra de usuários-alvo, sob a qual o ambiente é avaliado por meio de ferramentas de coleta e análise de dados.

A visão conceitual consiste em um *framework* composto por:

i) Uma taxonomia de percepção, composta por três dimensões principais, suas respectivas categorias de design e respectivos elementos ou mecanismos de suporte. Além disto, são descritas três dimensões adicionais que implicam diretamente as categorias de design e elementos de suporte: dimensões do papel/personagem (*persona*), fronteira (*boundary*) e temporal/histórica (*historical*);

ii) Um protocolo de planejamento da avaliação, que representa um instrumento de planejamento e execução do processo de avaliação. Este artefato auxilia na definição dos objetivos da avaliação, fatores a serem medidos, dimensões de percepção, fases do ciclo de vida em que a avaliação será aplicada, e assim por diante;

iii) Um conjunto de ferramentas de coleta e análise de dados, que engloba um conjunto de artefatos de apoio à condução da coleta e compilação dos dados obtidos pelas intervenções;

iv) Um conjunto de medidas de avaliação e escalas de qualidade do suporte à percepção, desenvolvidas para analisar os resultados obtidos por meio de instrumentos de avaliação e classificar o ambiente colaborativo ao nível de qualidade por meio da percepção dos participantes.

Por fim, a etapa de validação do modelo foi realizada em duas etapas. Primeiro, para melhorar o modelo de avaliação proposto, os artefatos do modelo foram expostos à apreciação de especialistas por meio da abordagem de painel de especialistas (BEECHAM *et al.*, 2005). Neste cenário, procurou-se expor nossos artefatos da taxonomia e modelo de avaliação ao escrutínio de especialistas para coletar a validade de critério e conteúdo do modelo. A revisão analisa os aspectos de utilidade, nomeadamente, clareza, relevância, consistência e completude dos itens do instrumento de medição.

Após esse refinamento, iniciou-se o planejamento e a execução de um conjunto de estudos de caso (WOHLIN *et al.*, 2012) por meio de uma avaliação em larga escala do nosso modelo de avaliação. Esta abordagem avalia a confiabilidade, validade e dimensionalidade do modelo proposto (TROCHIM; DONNELLY, 2001). Os dados foram reunidos como uma amostra única para análise de dados para cada estudo de caso. Em seguida, o modelo proposto foi avaliado quanto à confiabilidade e dimensionalidade. Considerou-se a consistência interna por meio do coeficiente alfa de Cronbach (CRONBACH, 1951) combinado com parâmetros da Teoria de Resposta ao Item (TRI) (PASQUALI; PRIMI, 2003; PASQUALI, 2020) para a medição da confiabilidade. A análise fatorial exploratória e a análise fatorial confirmatória foram aplicadas para testar a dimensionalidade (HAIR *et al.*, 2009; PASQUALI; PRIMI, 2003; IZQUIERDO; OLEA; ABAD, 2014).

RESULTADOS

Do ponto de vista do levantamento do mapeamento sistemático da literatura, identificou-se um conjunto de 92 elementos de suporte à percepção, que foram organizados em 17 categorias de design e cinco dimensões principais de percepção: contextual, de colaboração, situacional, de espaço de trabalho e histórica. O conjunto de elementos de suporte foram traduzidos em relação à classificação consolidada do framework 5W+1H.

Segundo, com base nos resultados do mapeamento sistemático, elaborou-se uma taxonomia de percepção que contempla elementos de suporte e aspectos de colaboração necessários para o trabalho cooperativo. Como resultado, construiu-se uma taxonomia multidimensional representada em três dimensões principais, nomeadamente, colaboração, espaço de trabalho e contextual, abrangendo 75 mecanismos de suporte descritos na literatura.

Terceiro, foi apresentado um novo método de avaliação do suporte à percepção centrado na perspectiva do participante, o qual incorpora um instrumento de medição baseado na Teoria de Resposta ao Item (PASQUALI; PRIMI, 2003; PASQUALI, 2020). A metodologia adotada permitiu construir e interpretar uma escala de qualidade para avaliar o nível de suporte à percepção considerando os 75 itens de avaliação identificados. Consequentemente, nós acreditamos que os aspectos essenciais do processo de colaboração são fornecidos mediante um suporte adequado. As correlações entre os elementos de design e o suporte à percepção fornecido foram definidas consoante à teoria e prática. Ainda, para utilizar adequadamente o método de avaliação proposto, foi projetado um processo de avaliação inspirado nas diretrizes de HCI e nas recomendações para avaliação da qualidade de produtos de software da norma ISO/IEC 25040:2011. Dessa forma, uma abordagem adaptativa foi desenhada, onde o examinador pode aplicar o modelo de avaliação completo ou selecionar as respectivas categorias de design e elementos de avaliação de interesse, ajustando assim os artefatos de coleta e análise de dados. Com a TRI é possível incluir novos itens na mesma escala de mensuração, como demográficos, usabilidade e UX, aumentando o potencial de avaliação.

O modelo de avaliação foi validado a partir das perspectivas do painel de especialistas e de estudos de caso. Primeiro, para melhorar o modelo proposto, expôs-se os artefatos do modelo à apreciação de especialistas por meio da abordagem de painel de especialistas (Capítulo 7) (BEECHAM *et al.*, 2005). Nesta etapa, os artefatos do modelo foram apresentados ao escrutínio de seis especialistas em Interação Humano-Computador (IHC) e sistemas colaborativos para avaliar o modelo face a validade de critério e de conteúdo, onde foram analisados aspectos de utilidade (*usefullness*) dos itens do instrumento de medição (clareza, relevância, consistência e completude). Os resultados sugerem que o modelo contempla os critérios desejados, dentre os quais, a aplicabilidade prática e a clareza nas definição dos artefatos do modelo e sua estrutura. Evidências indicam que os artefatos do modelo de avaliação (questões ou itens) são representativos, claros e relevantes, permitindo sua adoção em diversas situações; portanto, o modelo possui validade de conteúdo.

Após esse refinamento, iniciou-se o planejamento e a execução de três estudos de caso (WOHLIN *et al.*, 2012; YIN, 2009) por meio de uma avaliação em larga escala do modelo de avaliação para avaliar a confiabilidade do modelo e a validade do construto (TROCHIM; DONNELLY, 2001; DEVELLIS, 2016). Obteve-se a participação voluntária de 820 participantes, dentre os quais, 149 no primeiro estudo de caso, 422 no segundo e 249 no último.

O primeiro cenário de estudo de caso, descrito no Capítulo 8, foi desenhado para refinar o modelo e ajustar/adequar os artefatos. Neste cenário, avaliou-se os aspectos de colaboração proporcionados por um único ambiente colaborativo, considerando uma pequena amostra de participantes. Como resultado, foram encontrados indicadores adequados sob a perspectiva dos dados demográficos e da parametrização da TRI. Os resultados desta etapa de avaliação foram positivos. A escala de qualidade de percepção foi estabelecida segundo a TRI (PASQUALI; PRIMI, 2003; PASQUALI, 2020) e considerando a capacidade dos participantes de identificar informações de percepção; logo, pontuações mais altas indicam que os ambientes avaliados suportam facilmente mecanismos de percepção, enquanto os participantes com pontuações de habilidade mais baixas apresentam dificuldades em identificar mecanismos de percepção existentes.

No segundo cenário de estudo de caso, descrito no Capítulo 9, avaliou-se um conjunto de aplicações colaborativas para possibilitar a extração de resultados para cada um dos ambientes avaliados e assim verificar o comportamento do modelo de avaliação em cada um. Os resultados da avaliação da videoconferência foram positivos, e os ambientes mais familiares, sob o ponto de vista dos participantes, apresentaram o melhor desempenho. Neste cenário, as escalas de qualidade da percepção foram estabelecidas considerando a capacidade dos participantes em identificar informações de suporte. Na escala desenvolvida, pontuações mais altas indicam que os ambientes avaliados suportam facilmente mecanismos de percepção, isto é, participantes com pontuações de habilidade mais altas podem identificar adequadamente os mecanismos de percepção existentes. Neste cenário, foi aplicada a análise fatorial exploratória (do inglês, EFA) e a análise fatorial confirmatória (do inglês, CFA) para testar a dimensionalidade do modelo. Os resultados da EFA indicaram forte tendência ao modelo unidimensional (critério de raiz latente) (HAIR *et al.*, 2009), legitimando a correlação entre os itens de avaliação e o traço latente observado. Os resultados da CFA demonstram a validade de constructo do modelo: todos os fatores apresentaram confiabilidade composta (*composite reliability*) e cargas fatoriais adequadas (TAVAKOL; WETZEL, 2020; PASQUALI; PRIMI, 2003).

No terceiro cenário, descrito no Capítulo 10, os artefatos do modelo de avaliação foram apresentados à apreciação de diferentes examinadores para verificar a adequação do processo, suas atividades e artefatos relacionados. Este cenário envolveu a participação de 25 examinadores, divididos em 7 grupos de avaliação. Os resultados da avaliação de utilidade (*usefullness*) mostrou que, do ponto de vista do propósito (*purposefull*) e da determinação inequívoca (*unambiguous determination*), o modelo apresentou bons resultados: identificou-se considerável clareza e consistência nos artefatos e atividades de avaliação, permitindo sua aplicação sem maiores dificuldades pelos examinadores mesmo aqueles pouco familiarizados com a abordagem. Na perspectiva de aplicabilidade do modelo, que considera autenticidade, generalidade, usabilidade e completude, os examinadores demonstraram boa operabilidade em aplicar o modelo de referência, embora tenham sido supervisionados durante todo o processo. Quanto à completude, os examinadores indicaram que a representação contém todas as afirmações sobre o domínio, de forma correta e relevante.

Por fim, com base no compilado das 820 observações obtidas nos três cenários de avaliação, construiu-se uma escala global da percepção (descrita no Capítulo 11). Esta escala assume as estimativa das habilidades dos participantes e dos parâmetros dos itens como uma estratégia de equalização, conforme preconizado no método de estimativa multigrupo da TRI (*IRT multigroup estimation method*) (CHALMERS, 2012). Nesta abordagem, a calibração foi realizada por análise de máxima verossimilhança para dados politômicos (escala gradual de Samejima) (SAMEJIMA, 1969), usando a abordagem do algoritmo Metropolis-Hastings Robbins-Monro (MHRM) (CAI, 2010). Como resultado, a validação da escala global foi positiva, considerando as perspectivas de consistência interna, confiabilidade e dimensionalidade (EFA e CFA). Em síntese, a análise fatorial exploratória (do inglês, EFA) indicou forte tendência ao modelo unidimensional (critério da raiz latente) (HAIR *et al.*, 2009), legitimando a correlação entre os itens de avaliação e o traço latente observado; os resultados da análise fatorial confirmatória (do inglês, CFA) demonstram a validade de construto do modelo (FORNELL; LARCKER, 1981; HAIR *et al.*, 2009; TAVAKOL; WETZEL, 2020).

CONSIDERAÇÕES FINAIS

Acredita-se que a percepção está intrinsecamente relacionada às habilidades dos participantes para identificar, compreender e projetar suas ações. Assim, avaliar adequadamente o suporte de um ambiente colaborativo torna-se possível se considerarmos os elementos de suporte na perspectiva do participante. Assumindo um ambiente colaborativo plural, onde diferentes participantes com diferentes competências, conhecimentos e sabedoria se encontram e interagem, o modelo de avaliação do suporte à percepção desenvolvido constrói uma representação destes perfis existentes através de um amplo espectro de capacidades individuais.

A metodologia permitiu construir e interpretar uma escala de qualidade de percepção para avaliar o suporte à percepção na perspectiva do participante. Os cenários avaliados corroboram a validade do modelo. Primeiro, os resultados do painel de especialistas sugerem que o modelo possui validade de conteúdo. Segundo, os cenários de estudo de caso demonstraram: *i)* indicadores adequados do ponto de vista dos dados demográficos e da parametrização da TRI; *ii)* a aplicabilidade da escala de consciência na perspectiva do participante; e *iii)* a replicabilidade do modelo para diferentes cenários e contextos.

LIMITAÇÕES

Esta pesquisa assume a **transferibilidade** da taxonomia de percepção e a **generalização** do modelo de avaliação. A transferibilidade significa que o conhecimento gerado pela investigação não é generalizado; só é transferido para um contexto semelhante (transferindo conhecimento gerado de um contexto emissor para um contexto receptor) (POLIT; BECK, 2010). Assim, o leitor deve identificar o quanto o conhecimento se aplica a outro problema. Generalização é o grau em que os resultados podem ser aplicados a um contexto mais amplo. Os resultados da investigação são considerados generalizáveis quando as conclusões podem ser aplicadas à maioria dos contextos, à maioria das pessoas, na maior parte do tempo (POLIT; BECK, 2010).

Por fim, o modelo de avaliação foi desenvolvido com base na TRI, resultando numa abordagem de avaliação adaptativa/flexível; assim, novos itens de avaliação (ou seja, mecanismos de percepção ou outros aspectos (IHC) podem ser incluídos no modelo. É imprescindível observar que o modelo deve ser calibrado para os itens desejados sempre que novos itens forem incorporados. Além disso, a etapa de calibração do modelo TRI pode ser relativamente complexa se o examinador tiver conhecimento estatístico insuficiente. Ferramentas gerais de análise estatística, como R, são necessárias para executar os *scripts* de calibração e análise do modelo.

Palavras-chaves: Percepção; Sistemas Colaborativos; Modelo de avaliação.

ABSTRACT

[Context] We define awareness as a set of processes in which we recognize, organize, and make sense of the stimuli we receive from our environment. It has been a valuable concept in Collaborative Systems since its formation, and awareness research has followed the whole field's evolution over the last decades. We consider awareness the backbone of a collaborative environment; all collaborative concepts are archived through it. [Problem] According to the literature, awareness is a well-known but still not fully reached concept; this concept remains difficult to grasp, and researchers should achieve a better understanding of support by conceiving and testing novel technology that provides awareness. At this point, two issues remain open: first, how to understand awareness and how it should be faced in collaborative applications, and second, how to establish a basis for evaluating collaborative interfaces focused on awareness. [Goal] This work consists of establishing an awareness assessment model for collaborative interfaces from participants' perspectives by analyzing awareness information provided by the application. [Method] This research was carried out in four steps. First, we perform a systematic mapping study to identify the state-of-the-art strategies adopted in awareness development and evaluation and how they are used in the collaborative context. Second, we elaborate a taxonomy that contemplates awareness and collaboration aspects necessary for cooperative work. Third, based on the foundations of the literature, we elaborated an awareness assessment model. Finally, we validated the assessment model from two perspectives: *i)* to improve the proposed model, we expose the model's artifacts to the appreciation of six experts through the expert panel approach. *ii)* to verify the instrument's reliability and construct validity, we performed three case studies through a large-scale evaluation of the assessment model. In total, 25 examiners participated, and 820 user observations were collected. [Result] We performed a systematic mapping study to identify awareness support in the collaborative system context. We compiled the results into a multidimensional taxonomy of three main awareness dimensions: collaboration, workspace, and contextual, encompassing 75 awareness mechanisms described in the literature. Based on these results, 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 (IRT). [Conclusion] The methodology allowed us to construct and interpret an awareness quality scale to evaluate the awareness support from the participant's perspective. The scenarios evaluated corroborate the validity of the model. The expert panel results suggest that the model has content validity. The case study scenarios demonstrated: *i)* suitable indicators from the perspective of demographic data and IRT parameterization; *ii)* the applicability of the awareness scale over the participant's perspective; and *iii)* the model replicability for different scenarios and contexts.

Keywords: Awareness; Collaborative systems; Assessment model.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

3C	Communication, Coordination, and Cooperation
5W+1H	Who, What, Where, When, Why, and How
ACM	Association for Computing Machinery' Search Engine
BBB	BigBlueButton - Moodle
BSB	Backward Snowballing
CAT	Computer-based Adaptive Test strategy
CB	CounterBalanced design
CFA	Confirmatory Factor Analysis
CSCW	Computer-Supported Cooperative Work
CTT	Classical Test Theory
CR	Composite Reliability
DEF	Data Extraction Form
EC	Exclusion Criteria
EFA	Exploratory Factor Analysis
e.g.	"For Example" expression
EG	Equivalent Group design
FSB	Forward Snowballing
GQM	Goal Question Metric approach
HCI	Human-Computer Interaction
IC	Inclusion Criteria
i.e.	"That Is" expression
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers' Search Engine
IIC	Item Characteristic Curve
IIF	Item Information Function

IRT	Item Response Theory
ISO	International Organization for Standardization
MHRM	Metropolis-Hastings Robbins-Monro algorithm
PICO	Population, Intervention, Comparison, and Outcomes
SG	Single Group design
UX	User Experience

LIST OF SYMBOLS

$\ldots$	Ellipsis (continue in like manner)
AND	Logical operator AND
OR	Logical operator OR
$-$	Minus sign (subtraction operator)
$\#$	Item ID (item identification)
$>$	Strict inequality (greater than)
$<$	Strict inequality (less than)
$\forall$	For all (universal quantifier)
$\in$	Belong to (is an element of/set membership)
$\geq$	Inequality (greater than or equal to)
$\leq$	Inequality (less than or equal to)
$=$	Equal sign (equality)
$*$	Boundary probabilities
$()$	Parentheses (calculate expression inside first)
$\{\}$	Set (a collection of elements)
α	Cronbach's alpha coefficient
θ	Ability level
μ	Mean
$\mu(X)$	Mean for population X
$\mu(Y)$	Mean for population Y
$\mu(a_I)$, $\mu(a_J)$	Means of the discrimination parameter on scales I and/or J
$\mu(b_I)$, $\mu(b_J)$	Means of the difficulty parameter on scales I and/or J
$\mu(\theta(b_I))$, $\mu(\theta(b_J))$	Means of the difficulty parameter defined on scales I and/or J
σ	Standard deviation
$\sigma(X)$	Standard deviation for population X

$\sigma(Y)$	Standard deviation for population Y
$\sigma(b_I)$, $\sigma(b_J)$	Standard deviations over one or more items expressed on the scales I and/or J
$[\mu, \sigma]$	IRT Scale
$[\mu(X), \sigma(X)]$	IRT Scale for population X
$[\mu(Y), \sigma(Y)]$	IRT Scale for population Y
θ_j	Ability level (latent trait) of the j-th individual
$\tau_x(\theta_j)$	True score equating on test X which equivalent θ_j
$\tau_y(\theta_j)$	True score equating on test Y which equivalent θ_j
a	Discrimination value
A^*	Arbitrary constant A
a_i	Discrimination (or slope) value of item i, with a value proportional to the slope of the item's characteristic curve in the point b_i;
$a_{i1}, a_{i2}, ..., a_{im}$	The factor loading for the i-th test
a_{Ii}, a_{Ji}	Couples of discrimination for the item i on scales I and J
b	Difficulty value
B^*	Arbitrary constant B
b_i	Difficulty (or position) value of item i, measured on the same skill scale
$b_{i,k}$	Difficulty parameter of the k-th category of item i
b_{Ii}, b_{Ji}	Couples of difficulty for the item i on scales I and J
$\bar{c}$	Average inter-item covariance
c_i	Item parameter that represents the probability of individuals with a low ability to correctly answer the item i (it refers to the probability of a random hit)
d	Awareness dimension
e	Euler number (mathematical constant equals to $2.71828\ldots$)

e_i	Error specific to the i-th test that is uncorrelated with any common factors
f	Constant scale factor
$F_1, F_2, ..., F_m$	The common uncorrelated factors F
f_r	Observed score equating function
$f_r(x\|\theta_j)$	Distribution of correct scores over the first r items for examinees at the ability θ_j
g	Awareness goal
i	Measurement item
j	Participant (the examinee)
k	Score
m	Total of participants
n	Total of item scores
o	Total of measurement items
$\bar{v}$	Average variance
D	Set of Awareness dimensions
F	Set of Constant scale factors
G_d	Set of Awareness Goals of the awareness dimension d
I	Set of Measurement items
I_{gd}	Set of Measurement items related to the goal g of the awareness dimension d
J	Set of Participants
K	Set of Scores
$l_Y(x)$	Mean equating to transform a scale $[\mu(X), \sigma(X)]$ on test X to the scale $[\mu(Y), \sigma(Y)]$ of test Y
$m_Y(x)$	Mean that is equivalent to the score x in the test X to the test scale Y
N	Number of items in the test

p	Number of variables
P	Certain probability; Also refer to Population
$P1$, $P2$	Different populations (population P1; population P2)
Q	Question
X	Text X
X_i	The $i-th$ standardized score
$X1$, $X2$	Different sub-samples of X (sample X1; sample X2)
Y	Text Y
$Y1$, $Y2$	Different sub-samples of Y (sample Y1; sample Y2)
$P(\theta)$	Function of ability
$I(\theta)$	Item information function
$I_i(\theta)$	Amount of information for item i at ability level θ]
$I_{i,k}(\theta)$	Amount of information associated with a particular item response category k
$GI(\theta)$	Awareness goal's information function
$AI(\theta)$	Awareness dimension's information function
$P_i(\theta)$	Information provided by item i at the ability score θ
$P_{i,k}(\theta)$	Probability of an participant at the ability level θ chose a score k in the measurement item i
$P^*_{i,k}(\theta)$	Boundary probabilities of function $P_{ik}(\theta)$
$P^{*\prime}_{i,k}(\theta)$	Firs derivatives of function $P^*_{ik}(\theta)$
$P(U_{ij} = 1\|\theta_j)$	Probability of an individual j with ability θ_j correctly answering item i (also known as the item response function)
$Q_i(\theta)$	Probability of an incorrect response (do not select the category k as a response to assessment item i)
$U_{i,j}$	Dichotomous variable that takes the values 1, when the individual j correctly answers item i, or 0 otherwise

CONTENTS

1 INTRODUCTION

Needs like 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 are daily tools that allow people to connect, interact, and collaborate. A collaborative system, also known as Collaborative Software or Groupware, is 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 (ELLIS; GIBBS; REIN, 1991). These systems help members of a group work together and allow group members to share information and use that information to support working together (GEORGE, 2003).

Collaborative systems allow group members to communicate clearly with each other, coordinate their work, and cooperate. These three aspects are the pillars of the collaboration process, called the 3C collaboration model (communication, coordination, and cooperation). Collaboration occurs when two or more people, entities, or organizations work together to complete a task or achieve a goal (GEORGE, 2003). To provide the 3C collaboration model in collaborative systems, cues/information must be made available, allowing participants[1] to communicate, coordinate, and cooperate. This support involves a fundamental element of a collaborative approach: the awareness (DOURISH; BELLOTTI, 1992).

Awareness has been a significant concept in Collaborative Systems (TENENBERG; ROTH; SOCHA, 2016) and is an essential part of it (GROSS, 2013). Over the last three decades, different awareness types have emerged in the literature. The works of Seebach, Beck, and Pahlke (2011), Antunes et al. (2014), Gallardo, Bravo, and Molina (2018) and Mantau, Berkenbrock, and Berkenbrock (2017) present a broader list of awareness types. Detailed background on the origins of awareness, early ethnographic and technology studies that brought about the fundamental insights we found in (GROSS, 2013). A brief review of awareness support technologies used in collaborative systems is presented by Lopez and Guerrero (2017).

We consider awareness the backbone of a collaborative environment; all collaborative concepts are archived through it. 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. Second, each piece of awareness information supports the Cs of the 3C model. For example, workspace awareness

[1] We use the term **participant** to refer to the person who is participating, collaborating, or inserted in a collaborative context; The term **user** is adopted in relation to the person who is using a collaborative tool (computational application).

helps people move between individual and shared activities, provides a context for interpreting others' expressions, allows anticipating actions, and reduces the effort spent coordinating activities (GREENBERG, 1997).

In this sense, awareness is a process that occurs at three basic levels of abstraction: representation, understanding, and projection (BREZILLON et al., 2004).

At the representation level, we consider awareness through design mechanisms/elements that provide participants with cues about "what is going on"; These awareness mechanisms represent information regarding events and actions of all involved, whether individual, others, group, or the system itself. A solid base of elements/mechanisms is a constant concern throughout the area's evolution, and contributions can be seen in the works of (GUTWIN; GREENBERG; ROSEMAN, 1996; GUTWIN; GREENBERG, 2002; KIRSCH-PINHEIRO; DE LIMA; BORGES, 2003). In these works, awareness is related to the 5W+1H questions: who, what, where, when, why, and how.

To guarantee awareness aspects in a collaboration, it is necessary to provide a wide variety of information, such as identity, location, activity level, actions, intentions, modifications, objects, extensions, skills, the sphere of influence, and expectations (GUTWIN; GREENBERG; ROSEMAN, 1996). Over the last three decades, different awareness types have emerged in the literature. The works Seebach, Beck, and Pahlke (2011), Antunes et al. (2014), and Mantau, Berkenbrock, and Berkenbrock (2017) present a broader list of awareness types.

At the understanding level, we consider awareness as an understanding of other peoples' activities that provides a context for their own (DOURISH, 2004); this is a set of processes in which we recognize, organize, and find meaning for the stimuli we receive from the environment (STERNBERG; STERNBERG; MIO, 2012). From this perspective, awareness is the state of being conscious of something. We consider this level to be a mental state where, through the stimuli received from the environment and its experience and knowledge, the individual forms an understanding/consciousness of the situation and how he fits into this context. Finally, understanding/consciousness allows individuals to project their future actions and ensures collaboration.

These three awareness gears must be in tune for collaboration. In this sense, 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.

1.1 PROBLEM

Awareness is a well-known but still not fully reached concept in collaborative environments. According to Gross (2013), this concept remains difficult to grasp, and future research should achieve a better understanding of supporting and effortless coordination by conceiving and testing novel technology that guarantees awareness aspects. Furthermore, since the 1980s, there has been no consensus on the awareness issue and how to understand it (SCHMIDT; RANDALL, 2016).

At this point, two main potential issues remain open. The first relates to understanding awareness and how it should be faced in collaborative applications. The second is establishing a basis for evaluating collaborative interfaces focused on awareness.

1.1.1 The awareness problem

Awareness is a multi-factorial problem. First, to provide efficient mechanisms, collaborative systems require balancing between four major factors (MANTAU; BERKENBROCK; BERKENBROCK, 2014, 2017): *i)* guaranteeing awareness aspects to users; *ii)* avoiding problems arising from a human perspective, like overload, intrusiveness, privacy, information representation, understanding, projection of human actions through the interface, and cognitive load; *iii)* dealing with inherent devices restrictions and limitations; and *iv)* design issues, including usability, accessibility, communication, mobility, and navigation. Collaborative interfaces must be equally concerned with security, access control, joining of individual and group jobs, junction of synchronous and asynchronous jobs, and so on (GREENBERG, 1997). We must consider these issues in designing, developing, and evaluating collaborative systems.

Considering awareness at a high level, people can differ in their understandings, and individual awareness can change as their background and the stimuli received change. People have different abilities in representing, understanding, and projecting human actions through interface (MANTAU; BERKENBROCK; BERKENBROCK, 2017). Sociotechnical factors such as participants' motivation, knowledge, and goals influence interaction (CRUZ et al., 2012; MANTAU; BERKENBROCK; BERKENBROCK, 2017).

Second, each piece of awareness information supports each C of the 3C model. For example, workspace awareness helps people move between individual and shared activities, provides a context for interpreting others' expressions, allows anticipating actions, and reduces the effort spent coordinating activities (GREENBERG, 1997). Furthermore, the knowledge of shared workspace enables users to project their actions in the environment (e.g., to cooperate) and exchange information with others in the group (e.g., to communicate). These aspects make awareness a crucial element in collaborative applications and a fundamental challenge (as detailed in Section 3.4.2).

Gutwin and Greenberg (2002) developed a descriptive theory of awareness, organizing existing research through a conceptual framework focusing on workspace awareness. The framework comprises design elements addressing the questions of who, what, where, when, and how. Combined with the ”why”, these questions constitute the 5W+1H framework and represent the primary awareness information that collaborative systems must support. Few papers address awareness from a broad point of view (COLLAZOS et al., 2019). Most of these studies consider just a specific kind of awareness in their approaches (PROUZEAU; BEZERIANOS; CHAPUIS, 2018). Many papers consider awareness dissociated from communication, coordination, and cooperation. The link between supported awareness elements/information and their influence on collaboration remains hard to achieve. Furthermore, few studies present methods or processes that help to provide awareness aspects in collaborative systems.

1.1.2 The evaluation problem

Different strategies in the literature seek to help designers implement awareness mechanisms in collaborative application design and development stages. On the other hand, the most common approaches are projected to a specific context and do not focus on evaluating these mechanisms nor the provided support from the user's point of view. For example, Gallardo, Bravo, and Molina (2018) presented a framework for the descriptive specification of awareness support. They focused on multimodal user interfaces for collaborative activities, and a tool to help engineering implement the awareness in collaborative applications was provided; however, few clues toward evaluating these mechanisms have been presented. In recent efforts, they evolved this specification technique approach into a visual modeling language, and a software specification technique was found (BRAVO et al., 2023); however, how these elements are captured/appropriated by participants and how to evaluate it remains open.

As presented by Santos, Ferreira, and Prates (2012), there are many methods to evaluate collaborative systems, whether pre-existing, new/ad hoc or adapted. In general, adapted methods are commonly used, as most of these methods are already consolidated and allow, with some adaptations, to include the collaborative part of the systems (SANTOS; FERREIRA; PRATES, 2012).

On the other hand, considering specific awareness evaluation approaches, few studies present methods or processes that provide awareness aspects in collaborative systems (see Section 3.4.2). Many of them are based on a limited number of explorations, making it difficult to generalize the knowledge. The works of Steinmacher, Chaves, and Gerosa (2013) and Lopez and Guerrero (2017) investigated challenges related to providing awareness support during collaboration. Both papers highlight the lack of consolidated awareness assessment methods that allow collaborative applications to be assessed precisely from their perspective. Finding a good starting point in the literature can be challenging for beginners in awareness design (NIEMANTSVERDRIET et al., 2019). With a blank slate for each new application, designers must reinvent awareness from their experience of what it is, how it works, and how it is used in the task at hand (COLLAZOS et al., 2019).

1.2 RESEARCH QUESTION

How can we systematically evaluate collaborative applications considering the provided awareness mechanisms?

1.3 AIMS OF STUDY

The work consists of establishing an awareness assessment model[2] for collaborative interfaces from participants' perspectives by analyzing awareness information provided by the application. 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. As specific objectives, we can highlight:

i) Identify the state-of-the-art approaches (models, methodologies, or processes) adopted in design, development, and especially, evaluation of collaborative systems, addressing awareness and collaboration concepts (3C model);

ii) Identify the awareness elements involved in awareness support that collaborative interfaces must support and how we can conceptually represent it;

iii) Establish an awareness taxonomy that enables collaborative applications to achieve, through adequate provision of awareness information, collaboration aspects necessary for collaborative work;

iv) Elaborate an Awareness Assessment Model to evaluate collaborative applications from participants' perspective through accessing provided awareness support;

v) Establish a global awareness scale interchangeable[3] for different scenarios;

vi) Elaborate a set of assessment artifacts that guide the entire evaluation process, from data collection instruments, analysis tools, and construction and interpretation of awareness scales;

vii) Validate the Awareness Assessment Model in different scenarios.

---

[2] We use the term **model** to refer to the assessment model as a whole, both the conceptual view and the assessment model. The term **instrument** is adopted in relation to the model artifacts, e.g., data collection, analysis tools, and construction and interpretation awareness scales.

[3] Our global awareness scale can be used interchangeably in different scenarios due to the IRT equating strategy adopted. The IRT equating (see Section 2.6.8) is the statistical instrument used to compare different test scores from different forms when IRT models assemble tests, allowing the scores from both tests to be suitably used. Furthermore, IRT values are invariant to different groups of respondents as long as individuals in these groups have their skills measured on the same scale (ANDRADE; TAVARES; VALLE, 2000).

1.4 JUSTIFICATION

This thesis can be justified from four aspects: relevance, originality, complexity, and contributions.

1.4.1 Relevance

This work has theoretical and practical relevance. Theoretically, the focus of this study is the evaluation of support for awareness and collaboration in collaborative environments from the user's perspective. As we can see in the literature, few works present methods or processes that assist in providing awareness in groupware systems, and most common strategies focus on the design/development stages or are ad-hoc evaluation models. Additionally, there are no standardized tests for awareness assessment; Thus, measures must be established to assess awareness and identify the criteria for achieving awareness indicators.

The practical relevance is verified due to the importance of supporting awareness in collaborative environments. We conducted an in-depth investigation of the last ten years of publications from the CSCW field, where we noted that awareness and collaboration aspects remain the backbone of environments designed for collaboration. Furthermore, due awareness is intrinsically related to the participant's skills in identifying, understanding, and projecting their actions. Assessing awareness support is possible only if we consider the awareness elements from the participant's perspective. As participants' understanding differs, the awareness support scale must represent individuals with lower or higher abilities.

Due to its methodological characteristics, we used the Item Response Theory (IRT) to develop our assessment model. IRT can develop a scale to measure latent traits, such as awareness and collaboration support, based on a set of items, which can also be placed on the same scale as the latent trait. Furthermore, IRT can analyze the dimensionality and discrimination of a set of items, which, according to Bartolucci (2007), is impossible in any other known statistical approach. Classical Test Theory, also used to measure latent traits, is supported by the relationship between observed scores and true scores in a test. It makes the instrument dependent on the sample, which does not occur in item response theory, where the focus is not on the test as a whole, but the item and the generated scale are independent of the sample.

In short, IRT presents itself as a tool capable of measuring latent traits, highlighting the relationships between respondents and items, and being valid for broader situations as it does not directly depend on the group of respondents. IRT places items and respondents on the same scale, allowing comparisons between respondents and identifying the relationship between the participants' ability and the perceived awareness support. Developing the awareness assessment model produces a set of assessment items calibrated on the same scale.

Once the items are calibrated, it is possible to insert new items on this same scale, create a bank of items, and check the awareness support from another group of respondents. This is because the parameters of the items and respondents are invariant in IRT (NGUYEN et al., 2014).

1.4.2 Originality

Few models seek to evaluate awareness support from the user's perspective, largely ad hoc models that do not present standardized tests for evaluating awareness (COLLAZOS et al., 2019). In this sense, Prouzeau, Bezerianos, and Chapuis (2018) highlights that it remains necessary to establish measures to assess awareness and identify the criteria for achieving awareness indicators.

Existing models described in the literature to measure awareness support (as described in Section 2.8) centrally focus on the design and development process or provide design considerations to support interaction during the design process. In addition, because many of them are ad hoc approaches, the validity and reliability of the proposed instrument were not statistically checked. Furthermore, studies that use statistical techniques are based on Classical Test Theory (CTT). However, this work proposes using Item Response Theory (IRT) to measure the awareness support provided, ultimately providing a quality scale from the participant's perspective.

In this context, no tools, techniques, or instruments have been found to assess awareness through the IRT approach. This work differs from others presented in the literature as it uses advanced statistical techniques, as with IRT. Furthermore, this work demonstrates that IRT can contribute to assessment and collaborative environments, as it brings the assessment of awareness support closer to the individual point of view.

This research proposes an objective approach to awareness assessment from the user's perspective, which considers the last decade of research in awareness support, representing an unexplored view. Furthermore, the consolidation of awareness literature is discussed and modeled through item response theory, previously unpublished in the CSCW literature.

1.4.3 Complexity

Awareness is not fully achieved in collaborative environments, and providing awareness support remains challenging (GROSS, 2013). Therefore, there is a need to develop new techniques to achieve a better understanding of supporting awareness by conceiving and testing novel technology that provides awareness aspects. As we consider awareness an individual understanding of a particular environmental object or stimulus, measuring the quality of awareness support provided by the collaborative applications from the participant's viewpoint is necessary.

The natural complexity surrounding the awareness concept in the collaborative applications context added to the need to provide support to the participant's understanding/consciousness about awareness concepts (being aware of 'what is going on" in the shared environment), makes the construction of a scale of complex awareness support. Although highly beneficial for constructing the model and the awareness support scale, the use of IRT requires the researcher to have in-depth knowledge of the topic, which is highly complex. The software used in IRT requires specific knowledge and skills for its use and for interpreting the results generated.

1.4.4 Contributions

We can highlight the potential contributions resulting from this study:

i) The establishment of an awareness taxonomy that allows applications to achieve, through its adequate provision, the collaboration aspects necessary for collaborative work;

ii) The elaboration of an awareness assessment model to evaluate collaborative applications that helps to select which awareness aspects are relevant for collaborative system evaluation and how to access it;

iii) The elaboration of a global awareness scale that, based on the IRT assumptions (HOLLAND; DORANS; PETERSEN, 2006), allows the comparison of the results of different assessments using the generated global scale;

vi) The elaboration of an awareness assessment repository containing a full set of assessment artifacts necessary to guide the examiners through the evaluation process, from data collection instruments, analysis tools, and awareness scale construction and interpretation.

1.5 THESIS METHODOLOGY

The Design Science Research approach (DSR) (BICHLER, 2006; HEVNER, 2007) inspired the research method adopted. DSR helps to construct a new purposeful artifact to address a problem and evaluates its utility for solving problems of that type (VENABLE; PRIES-HEJE; BASKERVILLE, 2012) instead of explaining an existing reality or helping to make sense of it (IIVARI; VENABLE, 2009).

This research was carried out in four steps, as presented in Figure 1.

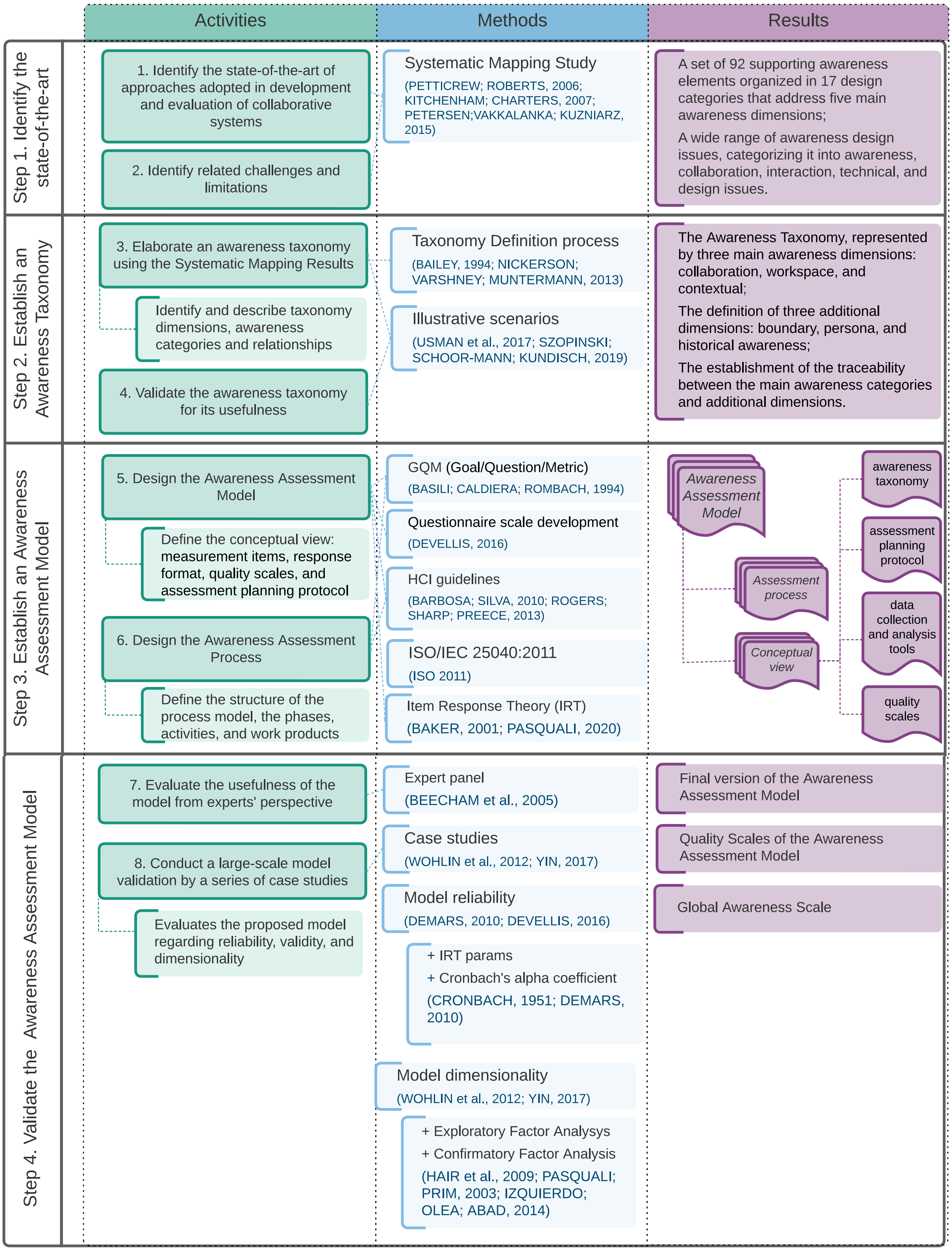


Figure 1 – Methodology overview

1.5.1 Step 1. Identify the state-of-the-art

In the first step, we perform a systematic mapping study to identify awareness support in the context of the collaborative system. We followed the guidelines in Petticrew and Roberts (2006), Kitchenham and Charters (2007) and Petersen, Vakkalanka, and Kuzniarz (2015). The analysis of the state of the art aims to identify the existing approaches adopted in developing and evaluating collaborative systems, address awareness and collaboration concepts (3C model), and identify related challenges and limitations. The systematic mapping process is divided into three phases: definition, execution, and analysis.

In the definition phase, research objectives are identified, and the systematic mapping protocol is defined. The protocol specifies the research questions and the procedures used to conduct the review, such as the definition of data sources, search query, inclusion/exclusion, and quality criteria. The execution phase consists of searching, identifying, and selecting relevant studies per the requirements defined in the protocol. During the analysis phase, we extracted the data using the data extraction form (DEF).

1.5.2 Step 2. Establish an Awareness Taxonomy

In the second step, we elaborate an awareness taxonomy that contemplates awareness elements and collaboration aspects necessary for cooperative work. Our Taxonomy Definition Method is based on Bailey's conceptual approach (BAILEY, 1994) and combines the guidelines presented by Nickerson, Varshney, and Muntermann (2013), Usman et al. (2017) and Szopinski, Schoormann, and Kundisch (2019). Our Taxonomy Definition Method consists of four phases: Planning, Identification, Design and Construction, and Validation (USMAN et al., 2017; SZOPINSKI; SCHOORMANN; KUNDISCH, 2019).

The planning phase defines the taxonomy's context and an 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 (NICKERSON; VARSHNEY; MUNTERMANN, 2013). Objective and subjective ending conditions are rules used to determine when to stop the interactive design and construction process.

Using our systematic mapping results, we collected data to define the new taxonomy in the identification phase. The terms were then collected, and redundancies and inconsistencies were identified and removed using a terminology control process.

We used a phenetic analysis during the design and construction steps, classifying elements by similarity (NICKERSON; VARSHNEY; MUNTERMANN, 2013). We identified different awareness characteristics/elements presented and clustered them into similar groups. 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 (USMAN et al., 2017). Illustrative scenarios and case studies are the two most used methods in the literature for the validation of taxonomies (USMAN et al., 2017; SZOPINSKI; SCHOORMANN; KUNDISCH, 2019) and are indicated when a conceptual approach is adopted (SZOPINSKI; SCHOORMANN; KUNDISCH, 2019).

1.5.3 Step 3. Establish a awareness assessment model

In the third step, based on systematic mapping results and awareness taxonomy, an awareness assessment model for collaborative environments is elaborated. Using the GQM (Goal Question Metric approach) (BASILI, 1992; VAN SOLINGEN; BERGHOUT, 1999), 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 (DEVELLIS, 2016). These elements represent the conceptual view of the assessment model.

Using the hierarchical structure in our awareness taxonomy, the evaluation objective is defined and systematically decomposed into factors to be measured. Elements are determined to support the development of the measurement instrument (questionnaire). The response format for the measurement instrument items was based on the response formats typically used in standardized questionnaires and assessment quality scales (BASILI, 1992; BAKER; KIM, 2017; DEVELLIS, 2016; PASQUALI, 2020).

We designed the awareness assessment process along with this stage, 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.

1.5.4 Step 4. Validate the Awareness Assessment Model

The validation of the proposed model was carried out in two stages. An overview of the validation process, detailing the steps and procedures, is presented in Part III, Chapter 6. The results of each step are presented in Chapters 7 to 10; Part IV, Chapter 11, describes the compilation of the global awareness scale.

Briefly, the model validation stage was performed as follows. First, to improve the proposed assessment model, we expose the model's artifacts to expert appreciation through the expert panel approach (BEECHAM et al., 2005). In this scenario, we seek to expose our taxonomy and assessment model artifacts to the scrutiny of experts to collect an accurate model criterion and content validity (detailed in Section 2.7). 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, the measurement instrument items' clarity, relevance, consistency, and completeness. The results of this step are presented in Chapter 7.

After this refinement, we started planning and executing a set of case studies (WOHLIN et al., 2012; YIN, 2009) through a large-scale evaluation of our assessment model. This approach evaluates the proposed model's reliability, validity, and dimensionality. We pooled the data as a single sample for data analysis for each case study. We evaluated the proposed model on reliability and dimensionality (detailed in Sections 2.7.1 and 2.7.2).

Data on reliability and construct validity were analyzed following the definition of (TROCHIM; DONNELLY, 2001) and the scale development guide (DEVELLIS, 2016). We considered internal consistency through the Cronbach alpha coefficient (CRONBACH, 1951) combined with IRT parameters for the reliability measurement. IRT allows us to evaluate the quality of the assessment items through θ, discrimination, and difficulty parameters. The test information function $I(\theta)$ was also used to calculate standard error and reliability (DEMARS, 2010).

Exploratory Factor Analysis (EFA) and Confirmatory factor analysis (CFA) were applied to test dimensionality (HAIR et al., 2009; PASQUALI; PRIMI, 2003; IZQUIERDO; OLEA; ABAD, 2014).

1.6 THESIS STRUCTURE OVERVIEW

This thesis is organized into four main parts. Part I presents the theoretical foundation and is composed of background (Chapter 2) and systematic mapping of literature chapters (Chapter 3). Part II describes the construction of the awareness assessment model and is composed of the awareness taxonomy (Chapter 4) and the assessment model (Chapter 5). Part III describes the assessment model validation process through an expert panel and case study scenarios, and is composed by Chapters 7 to 10. Part IV describes the global awareness scale compilation and is composed by Chapter 11. The discussions and conclusions are presented in Chapter 12. Supplementary materials are available in Appendices A to F.

Part I

Theoretical Foundations

2 BACKGROUND

Collaborative Systems are computer-based systems that support two or more people engaged in a common task or objective and provide an interface for a shared environment (ELLIS; GIBBS; REIN, 1991). These systems help members of a group work together and make it possible for group members to share information and to use that information to support working together (GEORGE, 2003).

This chapter presents the theoretical foundations used in this work to provide a better understanding of this research. We organized this chapter as follows. Section 2.1 also presents the concepts of collaborative systems. Section 2.2 contextualizes the 3C collaboration model. Section 2.3 presents an overview of what we understand about awareness and how this concept ensures collaboration occurs in collaborative environments. Section 2.4 provides a background understanding of taxonomy, mainly covering the taxonomy definition and assessment processes. Section 2.5 briefly presents the main assessment scales reported in the literature. Section 2.6 describes the statistical approach of the Item Response Theory, which we have adopted as a basis for developing our evaluation model. Sections 2.7.1 and 2.7.2 present the strategies adopted to validate our model concerning reliability and dimensionality properties. Section 2.8 presents an overview of related works found in the literature. Section 2.9 presents the chapter's discussions.

2.1 COLLABORATIVE SYSTEMS DEFINITIONS

In 1984, Irene Greif and Paul Cashman coined the CSCW (Computer-Supported Cooperative Work) for a workshop focused on understanding and supporting collaboration (SOEGAARD; DAM, 2012). Since its definition, the term has been used to refer to both the research field and collaborative applications/environments.

As a research field, collaborative systems (also called CSCW) are multidisciplinary (ELLIS; GIBBS; REIN, 1991) and integrate information processing and communication activities to help individuals work together as a group (PALMER; FIELDS, 1994). Olson and Olson (1997) refers to a research area that studies the use of computing and communication technologies to support group and organizational activity.

According to Gross (2013), the CSCW field aims to achieve a deep understanding of work and social interaction and to develop adequate concepts and tools for social interaction in groups and communities. Since its origins, CSCW has been a multidisciplinary field, using knowledge from different areas (i.e., psychology, anthropology, education) to investigate how groups of people relate and how they collaboratively develop their activities.

As a collaborative application, also called groupware, it was defined by Greif (1988) as "computer-assisted coordinated activity such as communication and problem-solving carried out by a group of collaborating individuals". Ellis, Gibbs, and Rein (1991) 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 is distinguished from normal software because it makes the user aware that he is part of a group, while standard software seeks to hide and protect users from each other (COLLAZOS et al., 2019).

CSCW is frequently used to refer to the research field (ELLIS; GIBBS; REIN, 1991; PALMER; FIELDS, 1994) and groupware (or collaborative system) to refer to the application or collaborative environment (GREIF, 1988; ELLIS; GIBBS; REIN, 1991). Some authors use collaborative systems to refer to the area and the environment (NICOLACI-DA-COSTA; PIMENTEL, 2011). We will adopt this second strategy and standardize our terminology.

2.2 THE COLLABORATION MODEL

Collaborative systems must support three fundamental pillars: communication, coordination, and cooperation, subsequently called the 3C collaboration model. The effective use of a collaborative system relies on promoting these characteristics (COLLAZOS et al., 2019), and it can be classified according to the support degree of each C. Furthermore, the cooperation pillar is a higher-level process, guaranteed by the two other: communication and coordination (GUTWIN; BARJAWI; PINELLE, 2016).

Figure 2 illustrates the 3C relationship. Fuks, Raposo, Gerosa, Pimental, et al. (2008) explore the 3C model to analyze and represent a collaborative application domain and serve as a basis for collaborative system development. Furthermore, the relationship among the Cs of the model may be used as a guide to analyzing a collaborative application domain (STEINMACHER; GEROSA; CHAVES, 2010). A variety of collaboration forms onto interrelationships between communication, coordination, and cooperation was presented by Fuks, Raposo, Gerosa, Pimentel, et al. (2008).

Communication is related to the exchange of information between members and is essential for the success of a collaborative system (FUKS; GEROSA; RAPOSO, 2002). Efficient communication allows group members to negotiate and decide about non-expected situations during the cooperation process (FUKS; RAPOSO; GEROSA; LUCENA, 2005).

Cooperation is the joint operation of the members of a group within a shared space that strives to accomplish tasks that are managed through coordination (FUKS; GEROSA; RAPOSO, 2002). Individuals cooperate by producing, manipulating, and organizing information, building and refining cooperation objects such as documents, spreadsheets, charts, etc. Cooperation is a higher-level process guaranteed by communication and coordination (GUTWIN; BARJAWI; PINELLE, 2016).

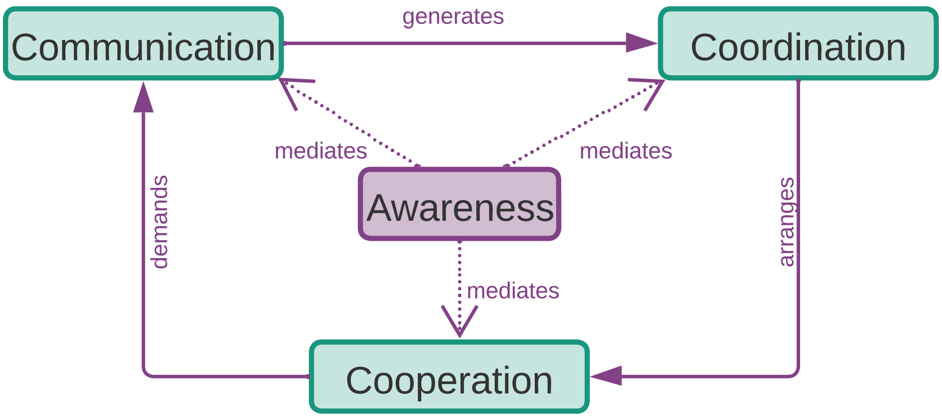


Figure 2 – 3C collaboration model relationship, adapted from Fuks, Raposo, Gerosa, Pimental, et al. (2008)

To ensure commitments and the realization of collaborative work through the sum of individual labor, it is necessary to coordinate the activities (FUKS; RAPOSO; GEROSA; PIMENTAL, et al., 2008). Coordination provides the effectiveness of communication and cooperation, and on the other hand, the lack of coordination generates conflicts in tasks or repetitive actions. It may be viewed as the link connecting communication and cooperation to enforce the success of collaboration (FUKS; RAPOSO; GEROSA; LUCENA, 2005).

The collaboration process occurs when a system supports each of these three dimensions. For this reason, a collaborative system developed focusing on a particular collaboration dimension (e.g., communication) must also support the other dimensions.

Although some authors use cooperation and collaboration interchangeably, we will consider the distinction noted by Molina et al. (2015) and Collazos et al. (2019). In this approach, collaboration is considered a superior activity in which, in addition to cooperating, the team members have to work together on daily tasks and towards a common outcome (MOLINA et al., 2015).

2.3 AWARENESS

An important aspect of collaborative environments is awareness, and it has become a standard point of consideration for interaction designers (NIEMANTSVERDRIET et al., 2019). This key concept can be defined as “the understanding of the activities of others involved and which provides a context for their activities” (DOURISH, 2004) or “set of processes in which we recognize, organize and make sense of the stimuli we receive from the environment we are in” (STERNBERG; MIO, 2006). In this context, awareness ensures that individual activities are tuned with the group’s objectives and enables collaboration. Without awareness, collective work is impossible; the group will be just an incoherent set of isolated pieces (BREZILLON et al., 2004). The awareness of individual and group tasks is critical to the success of the collaboration process (DOURISH; BELLOTTI, 1992).

Awareness elements[1] help people move between individual and shared activities, provide a context for interpreting the expressions of others, allow anticipation of actions, and reduce the effort expended to coordinate activities (GREENBERG, 1997). Awareness enables a given user to perceive the sensation of working in a group (COLLAZOS et al., 2019). Awareness is a concept that promises to improve the usability of collaborative applications (TANG; WINOTO; LEUNG, 2014); however, a clear general picture of awareness has not yet emerged from a collaborative perspective (LOPEZ; GUERRERO, 2017). For Gerosa, Fuks, and Lucena (2003), awareness information is essential to ensure coordination, cooperation, and communication (3C collaboration model). It allows 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 (GEROSA; FUKS; LUCENA, 2003).

Much research has addressed the study of awareness support within collaborative work, and various types of awareness have been proposed. The works of Seebach, Beck, and Pahlke (2011), Antunes et al. (2014), Gallardo, Bravo, and Molina (2018) and Mantau, Berkenbrock, and Berkenbrock (2017) present a broader list of awareness types; a brief review of awareness support technologies used in collaborative systems is presented by Lopez and Guerrero (2017).

On the other hand, awareness design is far from solving problems (BRAVO et al., 2023): *i)* it is difficult to find structured methods that allow designers to develop collaborative applications centered on awareness aspects (COLLAZOS et al., 2019); *ii)* there is a lack of established guidelines, specific tools, and systematic support for the modeling and design of awareness (BRAVO et al., 2023); and *iii)* research to conceive and test novel technology to support awareness and keep the coordination effort to a minimum are encouraged (GROSS, 2013).

A gap between theory and design remains an open area (RITTENBRUCH; MCEWAN, 2009). Consequently, the design of the awareness support has to be developed from scratch for each groupware system; designers must reinvent awareness from their experience of what it is, how it works, and how it is used in the task at hand (COLLAZOS et al., 2019). Thus, from the awareness support perspective, we need to fill the gap toward methods and tools to guide and facilitate the design, development, and, consequently, the assessment of the awareness support, aiding awareness designers

[1] We use the term **awareness element** to refer to the basic attributes or features related to or contribute to awareness. The term **awareness information** is adopted in relation to the data and content that informs users about the status, activities, and context within the collaborative system. The term **awareness mechanism** is adopted in relation to the processes, methods, and technologies used to gather, distribute, and present awareness information to users.

in building suitable awareness support for a groupware system according to the users' perspective (BRAVO et al., 2023).

2.4 TAXONOMY

Taxonomy describes and classifies existing or future objects of a domain (NICKERSON; VARSHNEY; MUNTERMANN, 2013). According to Bailey (1994), the term taxonomy can refer to both the process and the result; (i) as a process, taxonomy is the theoretical study of the classification, including its bases, principles, procedures, and rules; (ii) as a result, taxonomy is used for classification of empirical entities by a logical process derived from a conceptual or theoretical foundation. Nickerson, Varshney, and Muntermann (2013) define taxonomy as a set of dimensions, each consisting of "mutually exclusive" and "collectively exhaustive" characteristics. Mutually exclusive means that no object can fit more than one characteristic in a dimension; collectively exhaustive means that every object must have one characteristic in every dimension (OMAIR; ALTURKI, 2020).

2.4.1 Taxonomy assessment using illustrative scenarios

Taxonomy needs to be evaluated for its usefulness (NICKERSON; VARSHNEY; MUNTERMANN, 2013), and an illustrative scenario is the most common method to evaluate it in this context (PRAT; COMYN-WATTIAU; AKOKA, 2015). Illustrative scenarios apply the artifact in a real-world context or an artificial situation specially developed for evaluation purposes, intending to demonstrate the adequacy or usefulness of the artifact without temporal concern (e.g., use of the artifact for a while) (SZOPINSKI; SCHOORMANN; KUNDISCH, 2019).

According to Szopinski, Schoormann, and Kundisch (2019), the illustrative scenarios apply to two distinct types of objects: real/concrete objects or existing research. Using the real/concrete objects approach, the researcher evaluates the taxonomy by applying it to the objects for which the taxonomy is intended. After identifying the objects to be used, the researcher classifies them using the characteristics present in the taxonomy. Using the existing research objects approach, researchers assess taxonomy by applying it to research real-world objects. The assessment can be performed on a small (2 or 3) or large number of objects (50 or more) and conducted by a small group of researchers until the use of several independent researchers.

2.4.1.1 Applying in existing researches

In this approach, the taxonomy assessment seeks to identify its usefulness, effectiveness, clarity, understandability, and completeness. This is useful, as real-world objects help evaluate taxonomies and research about these objects (SZOPINSKI; SCHOORMANN; KUNDISCH, 2019).

The usefulness is related to the purposeful, unambiguous determination and applicability aspects (NICKERSON; VARSHNEY; MUNTERMANN, 2013). Purposeful is the significance and objectivity of taxonomy and its elements. Unambiguous determination is taxonomy's ability to represent its elements and characteristics concisely and unambiguously. Applicability seeks to assess its practical use for classifying, differentiating, and comparing objects.

In this context, the application of taxonomy allows us both to address whether a purposeful and unambiguous determination of taxonomy is possible by evaluating the practical applicability and to demonstrate whether a clear definition of taxonomy and its characteristics can be made (STRASSER, 2017). This approach also allows reflecting on the current state of research on an object (KHALILIJAFARABAD; HELFERT; GE, 2016), discovering similarities and differences between studies on this type of object (AGOGO; HESS, 2018), and identifying potential research gaps (HUMMEL; SCHACHT; MAEDCHE, 2016).

Strasser (2017) evaluated the usefulness in two steps. First, the taxonomy was applied to well-known publications, analyzing the purposeful and unambiguous determination aspects. Then, the practical applicability of the taxonomy was accessed through the selection of objects described in the literature and their classification using the constructed taxonomy. Khalilijafarabad, Helfert, and Ge (2016) developed their taxonomy through a literature review and evaluated by considering related publications (last two years).

To validate the results obtained, the taxonomy faced the literature to assess utility and robustness classification aspects. The evaluation stage sought to identify the most critical aspects/research described by the taxonomy and possible gaps and directions for further research. Agogo and Hess (2018) applied an existing taxonomy and analyzed how studied concepts can be differentiated using the selected taxonomy. Hummel, Schacht, and Maedche (2016) constructed their taxonomy through a literature review and evaluated by analyzing the frequency that respective aspects of the taxonomy are reported/used. This approach allows for identifying factors/characteristics that are most important/reported and potential research gaps.

2.5 ASSESSMENT SCALES

According to Stevens (1951), measurement is “the assignment of numbers to objects or events according to some rule”. Therefore, measurement consists of rules for assigning numbers to objects to represent the quantities of the attribute (CHADHA, 2009). In psychometry, a scale reminds us of a measuring instrument, a ruler, a thermometer, a weighing machine, or a scale that may be viewed as a set of elements, as in the Likert scale (CHADHA, 2009). In IRT, the scales aim to measure latent traits, opinions, or attitudes as objectively as possible. A scale consists of an arrangement, in the form of a graduated series of items, by which a characteristic is measured according to a previously determined number and is composed of categories, which are the gradations or alternatives of answers offered (GIL et al., 2002).

Stevens (1951) classified the different types of scales of measurement into four basic types, namely the nominal scale, the ordinal scale, the interval scale, and the ratio scale:

Nominal scale. It is the most basic scale of measurement. It consists of formulating a purposeful and homogeneous set of classes or categories of entities, facts, or data based on some trait, and assigning them some symbols or numerals as a way of differentiating two or more classes of entities or data, and keeping track of them (CHADHA, 2009). Specific numbers and symbols are simply labels, properties that define addition order or ratio are not achieved in nominal measurements (EMBRETSON, 1991).

Ordinal scale. In this scale, the entities or the data are ranked according to the degree to which they possess a particular attribute. Here, we seek to determine the rank order or inequality of elements to which numbers are assigned (CHADHA, 2009). Such relations are designated by the symbol $>$, which means “greater than” about particular attributes. Data from an ordinal scale satisfies all the properties of nominal data and adds numerical representation to an order of people or latent traits, thus allowing transformations and preserving both data distinctness and order (EMBRETSON, 1991).

Interval scale. In this scale, entities are not only ordered and ranked concerning some measured attribute but the distance or difference between neighboring ranks or states is also reflected, which is constant between each successive interval or rank. With the Interval Scale, we come to a quantitative form of measurement in the ordinary sense of the word (CHADHA, 2009). Statistical parameters are appropriate for interval scales, such as mean, standard deviation, and correlations (EMBRETSON, 1991).

Ratio scale. It is the most sophisticated of all the four measurement scales (CHADHA, 2009). Weight, length, time interval, electric resistance, and temperature measured on the Kelvin scale fall within the ratio scale measurement. A ratio scale has all the characteristics of nominal, ordinal, and interval scales and, in addition, an absolute or natural zero point representing the absence of magnitude of a variable attribute.

In short, it is possible to state that the appropriate scale for research must consider the nature of the variable to be measured, the respondent’s ability to make judgments, and the types of analysis to be developed.

2.6 ITEM RESPONSE THEORY

In many measurement situations, such as educational, social, or psychological, there is an underlying variable of interest, often something intuitively understood (BAKER, 2001). The development of appropriate scales to measure characteristics of individuals that cannot be measured directly has taken the attention of researchers (ANDRADE; TAVARES; VALLE, 2000).

Item Response Theory (IRT) is a set of mathematical models that seek to represent the probability of an individual giving a certain response to an item as a function of the item's parameters and the ability (or abilities) of the respondent (ANDRADE; TAVARES; VALLE, 2000). This relationship is always expressed so that the higher the skill, the greater the probability of hitting the item.

Item Response Theory is a latent trait theory applied primarily to tests of ability or performance. Latent trait theory refers to a family of mathematical models that relate observable variables (e.g. items in a test) and hypothetical unobservable traits or aptitudes (PASQUALI; PRIMI, 2003). In short, a stimulus (item) is presented to the subject, and he responds to it; the response the subject gives to the item depends on the level of the subject in the latent trait or aptitude (BAKER; KIM, 2017).

The usual approach to measuring ability is to develop a test consisting of several items (i.e., questions). Each of these items measures some facet of the particular ability of interest (BAKER; KIM, 2017). In IRT, the responses of a group of individuals are used to obtain estimates of the parameters of these items and the latent trait of each individual who responds to these items (AYALA, 2013b).

Applications of IRT are found in the most varied fields of research, like the quality of life (BORGES et al., 2017), the proficiency in specific knowledge (SCOTT; SCHUMAYER, 2015), the degree of depression (OLINO et al., 2012), the self-perceived cognitive functioning (RABIN et al., 2023), the environmental sustainability perception (VINCENZI et al., 2018), the effectiveness of organizational information and communication technology (ICT) (TRIERWEILLER et al., 2012), the supply chains performance (SANTOS; ARANTES, et al., 2023), the usability inspections (SCHMETTOW; VIETZE, 2008), the usability of sites (TEZZA; BORNIA; ANDRADE, 2011), the quality in e-commerce websites (TEZZA; BORNIA; ANDRADE; BARBETTA, 2018), the measuring physical, social, and self-presence in virtual reality environments (MAKRANSKY; LILLEHOLT; AABY, 2017), the chatbot conversation quality (SEDOC; UNGAR, 2020).

2.6.1 IRT characteristics

Human attributes, like intelligence, personality, attitude, skill, abilities, and so on, are likely to vary over a while, and sometimes even hours are sufficient to provide scope for such variations. Psychological attributes are highly dynamic and constantly undergo organization and reorganization (CHADHA, 2009). The IRT overcomes this aspect, whereas it allows matching items with the subject's aptitude. In other words, this allows more accessible items for subjects with lower abilities and more difficult items for more able subjects, producing comparable scores.

IRT was developed mainly to overcome the limitations of the Classic Test Theory, also known as latent trait theory (BORTOLOTTI et al., 2012). The IRT provides mathematical models for the latent traits, proposing forms representing the relationship between the probability of a respondent giving a certain response to an item and its latent trait and item characteristics (parameters) (BORTOLOTTI et al., 2012). Furthermore, in the IRT approach, person estimation is independent of the specific sample of the items, and item parameter estimation is not dependent on the particular sample of examinees (SCHUMACKER, 2005) – these properties are also called 'ìtem-free" and "person-free" estimation (AYALA, 2013a).

Item-free and person-free estimations are practical examples of the IRT invariance. Due to the skills/abilities and item parameters being estimated based on the responses of a group of individuals, once the skill measurement scale is established, the item parameters do not change. Consequently, IRT values are invariant to different groups of respondents as long as individuals in these groups have their skills measured on the same scale (ANDRADE; TAVARES; VALLE, 2000).

In other words, item parameters do not depend on the respondent's latent traits, and the individual's parameters do not depend on the items presented (BORTOLOTTI et al., 2012). The probability of responding to an item is precisely determined by the latent trait level of the respondent and not by their responses to other items in the group (FAYERS, 2004). So, IRT allows each item to be considered individually without revealing the total scores, so the conclusion does not depend on the instrument as a whole but on each item that comprises it (AYALA, 2013a).

With IRT, the respondents and items are on the same scale (metric), and the test is equally precise across the full range of possible test scores (HARVEY; HAMMER, 1999). The respondents are characterized by their position on the latent variable, and items are characterized according to their position and the ability to discriminate between the respondents (AYALA, 2013b). Responses given by respondents to a measuring instrument are used to estimate items and examinees simultaneously (on the same scale) (SCHUMACKER, 2005).

Other advantages include comparing latent traits of individuals in different populations when asked to complete tests or questionnaires with certain common items (ANDRADE; TAVARES; VALLE, 2000). It also allows individuals in the same population to be compared under totally different tests because *i)*item statistics are independent of the sample from which they were estimated, and *ii)* the respondents' scores are independent of the difficulty of the test (BORTOLOTTI et al., 2012). As a result, test analysis does not require rigorous parallel testing to assess reliability (SCHUMACKER, 2005).

2.6.2 IRT Models

IRT models show the relationship between an ability or trait θ measured by the instrument and the items' response. The item response may be dichotomous, such as right or wrong, yes or no, agree or disagree, or it may be polychotomous, such as a rating, scorer, Likert-type, or gradual-scale response on a survey (DEMARS, 2010).

In dichotomous models, there are only two categories; consequently, the model represents the probability θ of an individual j receiving the score of 1. Thus, the probability of a correct response is expressed as a function of $P(\theta_j)$. In this representation, the probability calculated for a specific value of θ can be interpreted as the probability of a correct response for an examinee randomly selected from a group of examinees with that value of θ (DEMARS, 2010).

2.6.2.1 IRT Dichotomous Models

The typical dichotomous models are represented by the logistic function of one, two, or three parameters (1PL, 2PL, and 3PL). Equation 1 presents the 3PL model formula (BAKER, 2001). The 1PL and 2PL IRT models are exceptional cases, or constrained versions, of the 3PL model. These models differ by the number of item parameters used in the function that models the relationship between θ_j and the item response (0 to 1).

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i)\frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \tag{1}$$

with,

$$i = 1, 2, ..., I; \quad j = 1, 2, ..., n; \quad k = 1, 2, ..., m; \quad b_{i,1} \leq b_{i,2} \leq ... \leq b_{i,mi}$$

where,

$P(U_{ij} = 1|\theta_j)$ is the probability of an individual j with ability θ_j correctly answering item i (also known as the item response function).

U_{ij} is a dichotomous variable that takes the values 1, when the individual j correctly answers item i, or 0 otherwise;

θ_j represents the ability (latent trait) of the j-th individual;

a_i is the discrimination (or slope) parameter of item i, with a value proportional to the slope of the item's characteristic curve in the point b_i;

b_i is the difficulty (or position) parameter of item i, measured on the same skill scale;

c_i is the item parameter that represents the probability of individuals with a low ability to correctly answer the item i (it refers to the probability of a random hit);

e is a mathematical constant, the Euler's number, approximately 2.718; and
i indicates the assessment item and j indicates the examinee.

The coefficient a_i is the discrimination parameter for item i, proportional to the slope of the item characteristic curve at point b_i. The parameter c represents the probability of a random hit. According to Baker (2001), the value of c does not vary depending on the skill level; thus, respondents with the lowest and highest ability have the same probability of a random hit.

Note that $P(U_{ij} = 1|\theta_j)$ represents the proportion of correct answers to item i among all individuals in the population with ability θ_j. This model is based on the fact that individuals with greater skill are more likely to get the item right and that this relationship is not linear (ANDRADE; TAVARES; VALLE, 2000). The skill scale is arbitrary; it is important to know the order of relationships existing between its points, not necessarily its magnitude. Parameter b is measured in the same unit as the skill and parameter c does not depend on the scale, as it is a probability and always takes values between 0 and 1.

The parameter b represents the skill required for a hit probability equal to $(1+c)/2$. therefore, the higher the value of b, the more difficult the item is, and vice versa. The parameter c represents the probability of an individual i with low ability θ answering the item correctly (also referred to as the chance probability).

The parameter a is proportional to the derivative of the curve's tangent at the inflection point (ANDRADE; TAVARES; VALLE, 2000); items with negative a are not expected under this model, as they would indicate that the probability of correctly answering the item decreases with increasing ability. Low a values indicate that the item has little discriminatory power (individuals with very different abilities have approximately the same probability of responding correctly to the item) (BAKER, 2001). Using the same logic, very high a values indicate items with very hard characteristic curves, which discriminate students into two groups: those with skills below the value of the b parameter and those with skills above the value of the parameter b.

2.6.2.2 IRT Polychotomous Models

Items with more than two categories are labeled polytomous or polychotomous (WEISS, 1995). These models are for items in which the categories are ordered (DEMARS, 2010). They are also appropriate for Likert-type items with an ordered response scale such as strongly disagree, disagree, neutral, agree, and strongly agree. The probability of scoring in or selecting each category is modeled in the polychotomous models. These models are for items in which the categories are ordered (BAKER; KIM, 2017).

Many IRT models have been developed to analyze ordinal polytomous items, such as the graded response model (SAMEJIMA, 1969), the rating scale model (ANDRICH, 1978), and the partial credit model (MASTERS, 1982). This study focuses on the graded response model because it is the most widely used IRT model for polytomous data (FORERO; MAYDEU-OLIVARES, 2009).

The Samejima (1969) graded response model assumes that the categories of responses given to an item can be ordered themselves. Thus, we can represent more information about individuals' responses than whether they got the item right, as in the dichotomous model.

In the gradual model, the category scores of an item i are ordered (from lowest to highest) and denoted by $k = 0, 1, ..., m_i$, where $m_i + 1$ is the number of categories of the i-th item. Therefore, the probability of an individual j choosing a particular category, or even higher, for item i can be given by extending the 3PL model with parameter $c = 0$ (see Equation 2).

$$P^+_{i,k}(\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_{i,K})}} \tag{2}$$

with,

$$i = 1, 2, ..., I; \quad j = 1, 2, ..., n; \quad k = 1, 2, ..., m; \quad b_{i,1} \leq b_{i,2} \leq ... \leq b_{i,mi}$$

where,

a_i is the discrimination parameter common to all categories of item i;
$b_{i,k}$ is the difficulty parameter of the k-th category of item i; and,
the other parameters are analogous to those defined in Equation 1.

As we assume the ordering between the level of difficulty for the categories of a given item i, we must have $b_{i,1} \leq b_{i,2} \leq ... \leq b_{i,mi}$. As a result, we can then find the probability of an individual j receiving a score k on item i by Equations 3 and 4.

$$P_{i,k}(\theta_j) = P^+_{i,k}(\theta_j) - P^+_{i,k+1}(\theta_j) \tag{3}$$

with,

$$P^+_{i,k}(\theta_j) = 1$$

$$P^+_{i,m_i+1}(\theta_j) = 0$$

thus,

$$P_{i,k}(\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_{i,k})}} - \frac{1}{1 + e^{-a_i(\theta_j - b_{i,k+1})}} \tag{4}$$

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 number equals $2.71828\ldots$);
m, n, and o are the total of participants, scores, and measurement items; and,
$b_{i,k}$ is the difficulty parameter of the k-th category of item i.

Samejima (1969) assume that an item i has $m_i + 1$ ordered categories, so we have m_i difficulty values. In this case, the number of parameters to be estimated for each item i will be given by the number of response categories m.

2.6.3 The Item Characteristic Curve

According to Baker and Kim (2017), the Item Response Theory model assumes that each examinee answers a test item considering some underlying ability, directly reflecting the response attributed to each test item. Thus, each assigned response to the item can be traced to a given participant's ability level. This ability score is denoted by the Greek letter theta (θ).

In IRT, each examinee can be labeled with a numerical value, a score, that places the examinee somewhere on the ability scale, and, at each ability level, there will be a certain probability that an examinee with that ability will give a correct answer to the item (BAKER, 2001). This probability will be denoted by $P(\theta)$.

The IRT model is based on the fact that individuals with greater skill are more likely to get the item right and that this relationship is not linear; the skill scale is an arbitrary scale where the importance is the existing order relationships between its points and not necessarily its magnitude. The probability of a correct response starts at zero at the lowest levels of ability and increases until the highest.

The IRT makes two basic postulates (PASQUALI; PRIMI, 2003):

- The subject's performance on a task (item of a test) can be predicted from a set of factors or hypothetical variables, said aptitudes or latent traits;
- The relationship between performance and latent traits can be described by an increasing monotonic mathematical equation called the Item Characteristic Curve (ICC).

The S-shaped curve, presented in Figure 3, describes the relationship between the probability of a correct response to an item and the ability scale (BAKER; KIM, 2017). As the theta increases, the probability of getting the item right also increases, so there is an increasing monotonic relationship between aptitude and probability of getting the item right (PASQUALI; PRIMI, 2003).

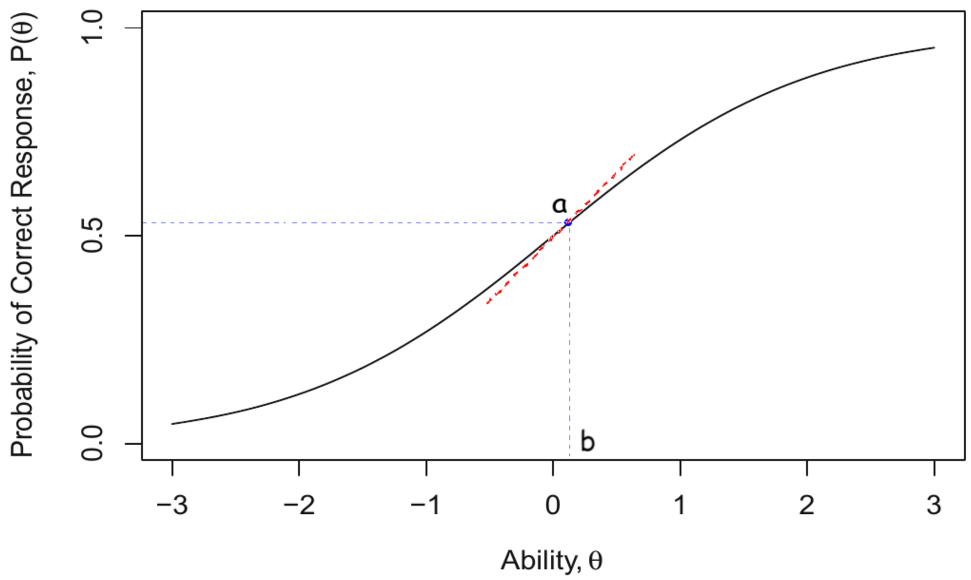


Figure 3 – The Item Characteristic Curve

2.6.4 Item Difficulty and Item Discrimination

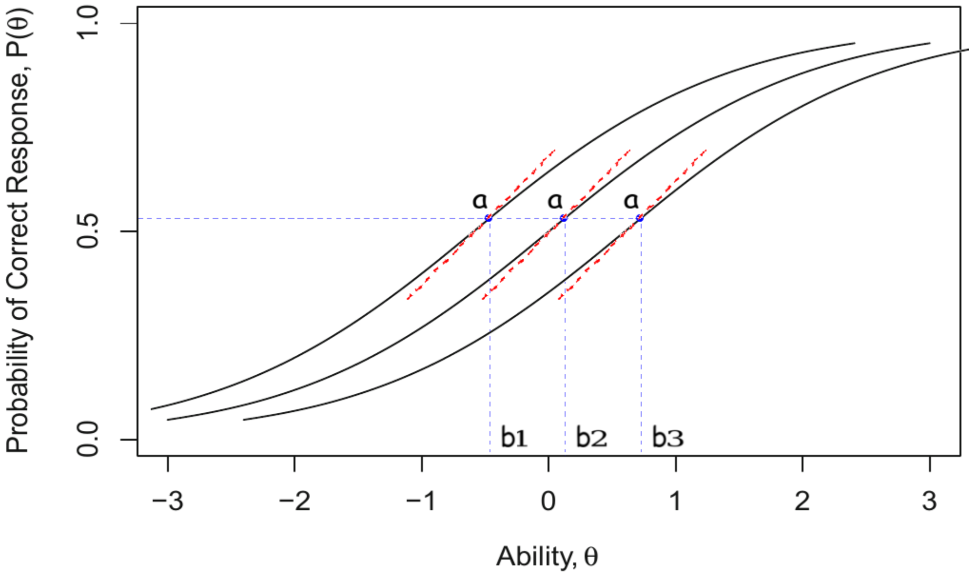
The difficulty is defined by the likelihood of the correct response, not the perceived difficulty or amount of effort required (DEMARS, 2010). The difficulty index (b) is the same metric as the proficiencies or traits. This metric is arbitrary, but is often anchored so that the proficiency distribution in a designated group has a mean of 0 and standard deviation of 1 (DEMARS, 2010).

Higher discrimination means that the item differentiates (discriminates) between examinees with different levels of the construct; thus, high discrimination is desirable (DEMARS, 2010). The instrument aims to differentiate (discriminate) between examinees who know the material tested and those who do not, or on an attitude scale, between positive and negative attitudes.

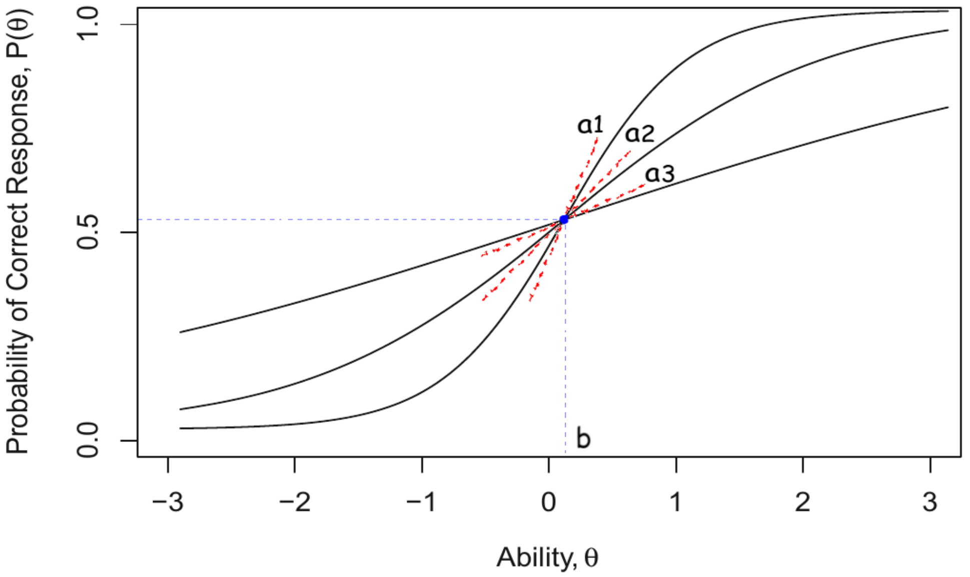
The model is based on the fact that individuals with greater skill are likelier to hit the item, and this relationship is not linear. This can be observed through the graph in Figure 4, where the ICC has the S-shape form (represented by the sigmoid function), whose inclination and displacement on the scale are defined by the item's parameters.

Figure 4 exemplifies a sigmoid function of the ICC. In the graph in Figure 4a, we see three distinct curves with identical discrimination values (a) but with different levels of difficulty (b). In this case, $b1$ represents an item with slightly lower difficulty, and $b3$ is an item with more incredible difficulty about parameter $b2$.

In contrast, in the graph in Figure 4b, we see three distinct curves with different discrimination values ($a1$, $a2$, and $a3$) but with the same level of difficulty (b). In this case, $a1$ represents an item with slightly less discrimination, and $a3$ is an item with a greater discrimination than parameter $a2$.



(a) same item discrimination; different item difficulty



(b) different item discrimination; same item difficulty

Figure 4 – The Item Difficulty and Item Discrimination

According to Baker (2001), an item measures ability with the greatest precision at the ability level corresponding to the item's difficulty parameter. The amount of item information decreases as the ability level departs from the item difficulty and approaches zero at the extremes of the ability scale. DeMars (2010) highlights that the accuracy of estimating θ increases with the number of items (as well as the nearness of the item difficulty to the y location), and the number of items necessary to estimate θ can vary depending on the match of the a and b. Shortly, better items will be more discriminating and helpful in estimating.

2.6.5 The Item Information Function

A measure widely used in conjunction with the item characteristic curve is the item information function. The Item Information Function (IIF) allows you to analyze how much information an item contains about the skill measure (ANDRADE; TAVARES; VALLE, 2000; BAKER; KIM, 2004). It indicates the amount of information an item presents within the skill scale. In the Samejima's gradual scale (SAMEJIMA, 1969), the item information function is defined as Equation 5 (SAMEJIMA, 1969).

$$I_i(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)} = \sum_{k=1}^{m_i} \frac{[P_{i,k-1}'(\theta) - P_{i,k}'(\theta)]^2}{P_{i,k}^*(\theta)} \tag{5}$$

with,

$$I_i(\theta) = \sum_{k=1}^{m_i} I_{ik}(\theta)P_{ik}(\theta)$$

$$P_i(\theta) = P(X_{ij} = 1|\theta)$$

$$Q_i(\theta) = 1 - P_i(\theta)$$

where,

$P_i(\theta)$ is the information provided by item i at the ability score θ;

$Q_i(\theta)$ represents the probability of an incorrect response (do not select the category k as a response to assessment item i);

$P_{ik}(\theta)$ represents the probability to select category k as a response to item i;

$I_{ik}(\theta)$ denote the amount of information associated with a particular item response category k;

$*$ represents the boundary probabilities [2].

Equation 5 shows the importance of the three parameters on the amount of information in the item. That is, the information is greater: *i)* when b_i approaches θ; *ii)* the larger the a_i; and, *iii)* the closer c_i gets to 0. The item information function can be graphically represented by curves that behave as shown in Figure 5.

In this example, there are 4 curves (P1 to P4). P1 is the one that presents the least information, being representative at ability score interval $[-4, -2]$; P2 and P3, respectively, present more information at intervals $[-2, 0]$ and $[0, +2]$; and P4 presents more information over skill region $[+2, +4]$.

[2] Both the item response category information function and the item information function can be expressed in terms of the boundary probabilities $P_{ik}(\theta)$, i.e., the interval in the ability scale where the participant is most likely to choose category k for item i (SAMEJIMA, 1969).

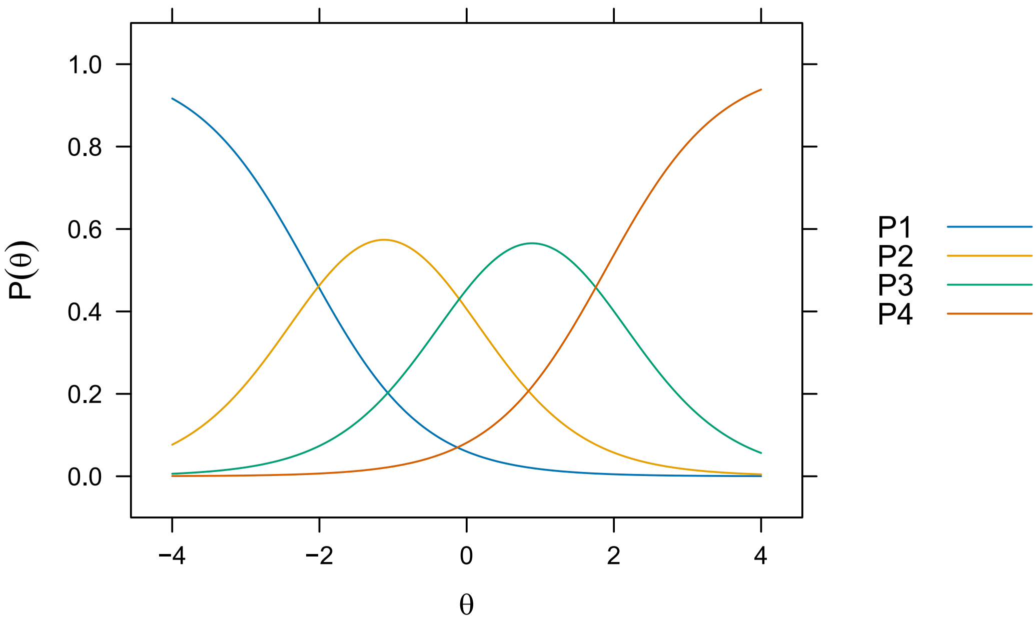


Figure 5 – Item Information Function example

According to Samejima (1969), the amount of information due to an item response category $I_{ik}(\theta)$ is a measure of how well responses in that category estimate the examinee's ability and the information share of a response category $I_{ik}(\theta)P_{ik}(\theta)$ is the amount of information contributed by the category to the item information.

2.6.6 The Test Information Function and standard error

An IRT test is used to estimate an examinee's ability from the perspective of a certain latent trait. We can obtain the amount of information yielded by the test at any ability level by applying the test information function (see Equation 6).

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta) \tag{6}$$

where,

$I(\theta)$ is the amount of test information at an ability level of θ;
$I_i(\theta)$ is the amount of information for item i at ability level θ; and,
N is the number of items in the test.

A test is a set of items; therefore, the test information at a given ability level is the sum of the item information at that level. The test information function will be much higher than that for a single-item information function because a test measures ability more precisely than a single item; thus, the more items in the test, the greater the information (BAKER, 2001). Consequently, more extended tests will measure an examinee's ability with greater precision than shorter tests.

In this classical approach, the standard error of measurement, or standard error of the estimated score, SE, is then based on the definition that observed score variance is calculated by true score variance plus error variance (DEMARS, 2010). Furthermore, a single estimate of the standard error is calculated regardless of the value of the observed score, X. With this definition, extreme scores (scores close to 0 or the maximum value) have more minor standard errors than middle scores. This definition does not require an estimate of reliability to calculate the standard error. Instead, the standard error is calculated first and then can be used to form a reliability-like coefficient (FELDT; CHARTER, 2003).

The information function can also be defined at the item level, and the item information functions sum to the test information function. Suppose a reliability estimate of the IRT scores is desired for a sample or population of examinees. In that case, it can be estimated based on the item parameters and the trait distribution in that group of examinees (DEMARS, 2010).

Figure 6 exemplifies the relationship between test information $I(\theta)$ and the standard error $SE(\theta)$ in the IRT. We can see that in locations where there is more information, the standard error is lower. The blue line represents the test information function; the red dotted line represents the standard error; the intersection point at the vertical red dotted lines represents the limits at which the model is more representative.

2.6.7 Test calibration

In the IRT model, one of the most important steps is estimating items' parameters and participants' skills, which are generally unknown, considering that only the participants' responses are available. In IRT, this estimation process is known as calibration (ANDRADE; TAVARES; VALLE, 2000).

From a theoretical viewpoint, the calibration can be divided into three situations: *i)* Knowing the items' parameters, only the skills can be estimated; *ii)* Knowing the participants' skills, the items' parameters are estimated; and *iii)* Items' parameters and respondents' skills are estimated simultaneously (this is the most common situation).

Several methods are used to estimate the parameters of test items, which depend on the number of groups and types of tests. Andrade, Tavares, and Valle (2000) describes six possible basic cases regarding the number of groups and types of evidence involved (see Figure 7). This configuration is valid for cases using two tests and two populations; however, situations involving more tests and/or populations are analogous.

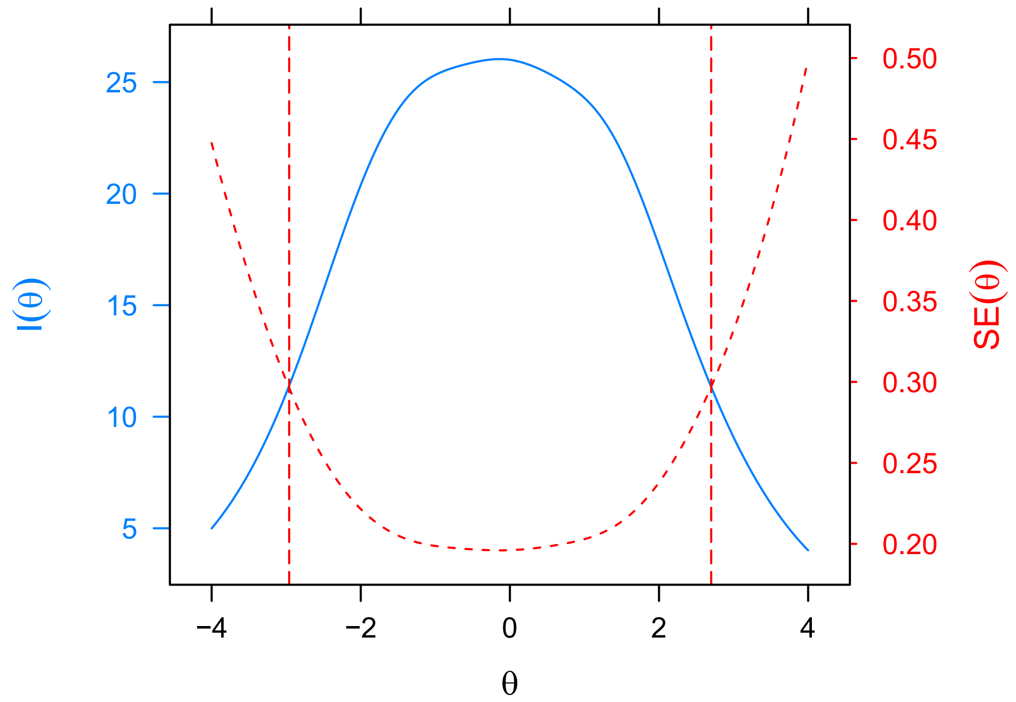


Figure 6 – Test information $I(\theta)$ and Standard Error $SE(\theta)$

Case a: a single group taking a single test. This is the most basic case, where all participants respond to the same assessment instrument, and consequently, the estimation corresponds directly to the population.

Case b: a single group divided into two subgroups, taking two completely different tests (no common items). In this scenario, it is enough for all items from both tests to be calibrated simultaneously. According to Andrade, Tavares, and Valle (2000), the fact that all individuals represent a random sample from the same population guarantees that all parameters involved will be on the same scale.

Case c: a single group, divided into two subgroups, taking two tests, only partially distinct, that is, with some common items. Via population, this case is similar to case b; via common items perspective, it resembles case f.

Case d: two groups doing a single test. In this scenario, as both populations take the same test, it is enough for the items to be calibrated using the responses of respondents from both groups simultaneously.

Case e: two groups taking two completely different tests (no common items). This is the only scenario in which we cannot use IRT. In this case, it is possible to calibrate the items of each test separately; however, no comparison can be made between the results observed in the two different populations.

Case f: two groups taking two tests, only partially different, that is, with some common items. In this scenario, the common items between different tests allow all

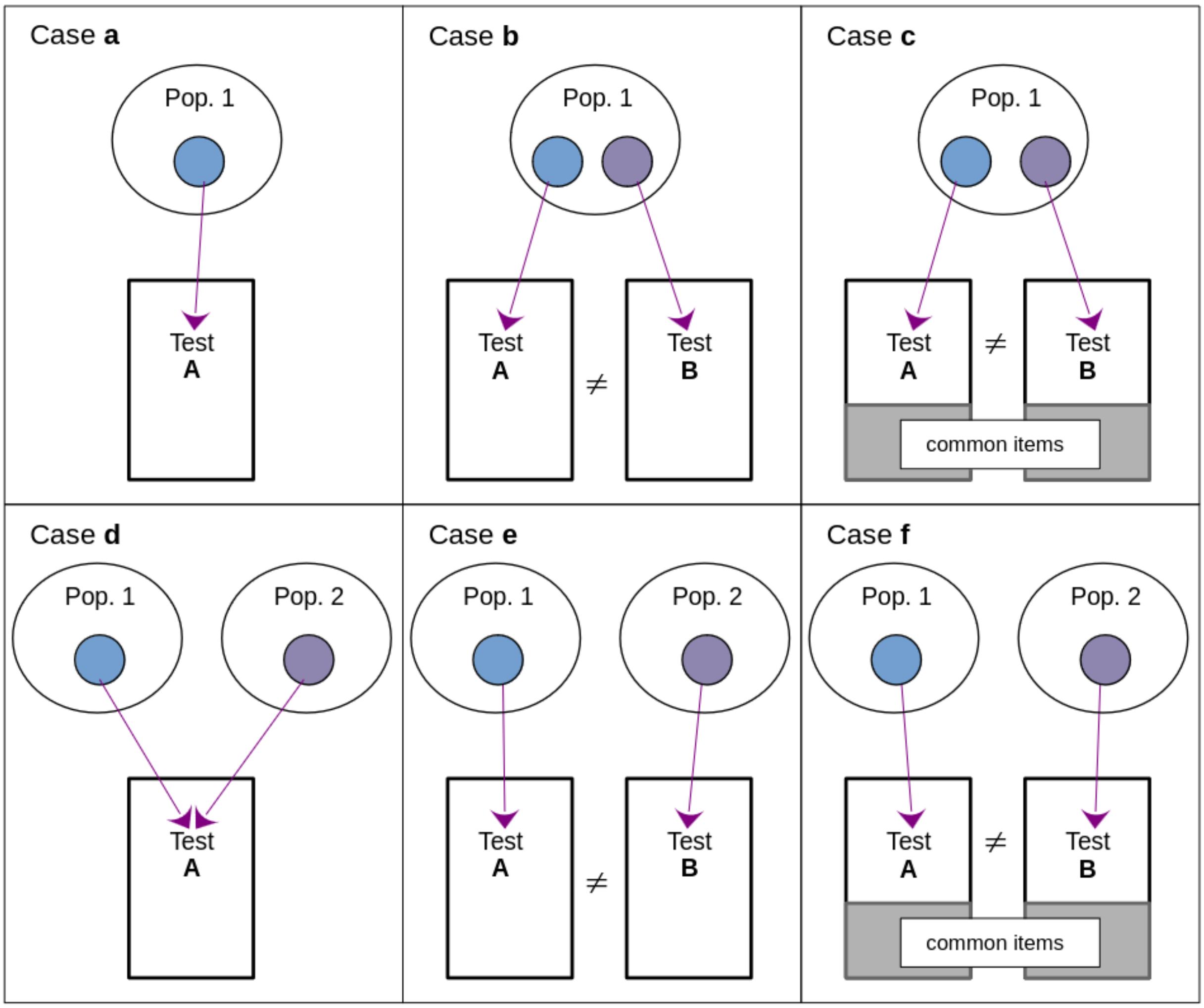


Figure 7 – Different tests scenarios
Adapted from Andrade, Tavares, and Valle (2000)

parameters to be on the same scale at the end of the estimation processes. As there is a set of items that connect the different populations, it is possible to make comparisons and construct a global scale.

2.6.8 Test equating

To estimate the participants' skills, it is necessary to equalize the items' parameters, that is, to make comparisons or place the parameters from different tests or skills of respondents from different populations on the same metric or common scale (ANDRADE; TAVARES; VALLE, 2000). The goal of equating is to produce a linkage between scores on two test forms such that the scores from each test form can be used as if they had come from the same test (DORANS; MOSES; EIGNOR, 2010).

IRT equating is the statistical instrument used to compare different test scores from different forms when IRT models assemble tests (SANSIVIERI; WIBERG; MATTEUCCI, 2017). The IRT equating allows the scores from both tests to be used interchangeably. According to Holland, Dorans, and Petersen (2006), five requirements are widely considered necessary for linking different test scores:

i) The equal construct requirement. The two tests should both be measures of the same construct (latent trait, skill, ability);

ii) The equal reliability requirement. The two tests should have the same level of reliability;

iii) The symmetry requirement. The equating transformation for mapping the scores of test Y to those of test X should be the inverse of the equating transformation for mapping the scores of test X to those of test Y;

iv) The equity requirement. It should be a matter of indifference to an examinee as to which of two tests the examinee takes and,

v) The population invariance requirement. The equating function used to link the scores of test X and test Y should be the same regardless of the choice of (sub) population from which it is derived.

2.6.8.1 Data collection for equating

According to Holland, Dorans, and Petersen (2006), numerous data collection designs have been used for score equating. Three basic approaches allow us to equalize the scenarios described in section 2.6.7: the single group (SG) design, the equivalent groups (EG) design, and the Counterbalanced (CB) design.

In the single group design, all examinees in a single sample of examinees from population P take both tests. The single group design is the earliest data collection design used for test linking, and it can provide accurate equating results from relatively small sample sizes (HOLLAND; DORANS; PETERSEN, 2006).

In the equivalent groups (EG) design, two equivalent samples are taken from a common population P; one is tested with X, and the other with Y (HOLLAND; DORANS; PETERSEN, 2006). EG design is performed in two ways: *i)* to take two random samples from P ($P1$ and $P2$) and test each with a single test; or *ii)* to construct two samples by balancing the X and Y tests.

In the counterbalanced (CB) design, the sample tests (X and Y) are randomly divided in half and, in each sub-sample ($X1, X2$ and $Y1, Y2$), the tests are taken in different orders (HOLLAND; DORANS; PETERSEN, 2006). The CB design contains the SG and EG designs: SG designs for $X1$ and $Y2$ and $X2$ and $Y1$; EG design for $X1$ and $Y1$ and $X2$ and $Y2$.

2.6.8.2 Equating scores

Many procedures have been developed for equating tests. Holland, Dorans, and Petersen (2006) considered three factors of the equating methods: *i)* by common population or common item; *ii)* by linear or nonlinear methods, and *iii)* by observed score or true score procedures.

i) Equating by common population or common item.

In equalization via population, we assume if a single group of respondents is subjected to different tests, it is enough for all items to be calibrated together to guarantee that they will all be on the same metric. In equalization by common elements, the guarantee that the involved populations have their parameters on a single scale will be provided by common elements between the populations, which will link them (ANDRADE; TAVARES; VALLE, 2000).

ii) Equating by linear methods

In linear equating, we assume that only the means and standard deviations of the scores on the two forms are equal (HOLLAND; DORANS; PETERSEN, 2006). Assuming the IRT scale $[\mu, \sigma]$ and the EG data collection design, we define the population means on X and Y, respectively $\mu(X)$ and $\mu(Y)$, and the standard deviations as $\sigma(X)$ and $\sigma(Y)$ Sansivieri, Wiberg, and Matteucci (2017). Then, we calculate the linear equating by Equation 7 (SANSIVIERI; WIBERG; MATTEUCCI, 2017).

$$l_Y(x) = y = \frac{\sigma(Y)}{\sigma(X)}x + [\mu(Y) - \frac{\sigma(Y)}{\sigma(X)}\mu(X)] \tag{7}$$

where,

$l_Y(x)$ indicates the mean equating to transform a scale $[\mu(X), \sigma(X)]$ on test X to the scale $[\mu(Y), \sigma(Y)]$ of test Y.

Similarly, we can also equate the means of the scores on X and Y by Equation 8 (SANSIVIERI; WIBERG; MATTEUCCI, 2017).

$$m_Y(x) = y = x - \mu(X) + \mu(Y) \tag{8}$$

where,

$m_Y(x)$ are the mean that is equivalent to the score x in the test X to the test scale Y.

iii) Equating by observed score or true score

For IRT equating, we consider two IRT assumptions: an examinee parameter is constructed as a unidimensional latent trait, and the examinee responses to the items are statistically independent (local independence of items) (SANSIVIERI; WIBERG; MATTEUCCI, 2017).

The observed score and the true score are statistical instruments used to compare different test scores from different IRT tests (LORD, 1980). Let scales I and J that linearly differ and θ_{Ij} and θ_{Jj} be values of the ability θ for examinee j on these two scales. The relationships between the item parameters can be formulated as Equation 9 (SANSIVIERI; WIBERG; MATTEUCCI, 2017).

$$\theta_{Jj} = A^{*}\theta_{Ii} + B^{*} \tag{9}$$

with,

$$a_{Ji} = \frac{a_{Ii}}{A^{*}}; \qquad b_{Ji} = A^{*}b_{Ii} + B^{*}$$

then,

$$A^{*} = \frac{\sigma(b_J)}{\sigma(b_I)} = \frac{\mu(a_I)}{\mu(a_J)} = \frac{\sigma(\theta(b_J))}{\sigma(\theta(b_I))}$$

$$B^{*} = \mu(b_J) - A^{*}\mu(b_I) = \mu(\theta_J) - A^{*}\mu(\theta_I)$$

where,

A^{*} and B^{*} are two arbitrary constants;

a_{Ii}, a_{Ji}, b_{Ii} and b_{Ji} are the couples of discrimination and difficulty for the item i on scales I and J;

$\mu(a_I)$, $\mu(a_J)$, $\mu(b_I))$, and $\mu(b_J)$ is the means of the a and b parameters on scales I and J;

$\sigma(b_I))$, and $\sigma(b_J)$ are the standard deviations over one or more items having parameters that are expressed on the two considered scales;

$\sigma(\theta(b_I))$ and $\sigma(\theta(b_J))$ are the standard deviations defined over two or more examinees having parameters that are expressed on the two defined scales;

The true scores equating on test X and Y, which equivalent θ_j, are defined respectively as Equations 10 and 11 (KOLEN; BRENNAN, 2014).

$$\tau_x(\theta_j) = \sum_{i:X} p_{ji}(\theta_j, a_i, b_i) \tag{10}$$

and,

$$\tau_Y(\theta_j) = \sum_{i:X} p_{ji}(\theta_j, a_i, b_i) \tag{11}$$

where,

θ_j the ability parameter for examinee j;
a_i the discrimination parameter for item i;
b_i the difficulty for item i;
p_{ji} as the probability of a response for examinee j to item i.

The observed score equating for test X and Y can be calculated as presented by Kolen and Brennan (2014). As defined by LORD (1980), the observed score equating can be reached by Equation 12.

$$\begin{aligned} f_r(x|\theta_j) &= f_{r-1}(x|\theta_j)(1-p_{jr}), \qquad x = 0; \\ &= f_{r-1}(x|\theta_j)(1-p_{jr}) + f_{r-1}(x-1|\theta_j)p_{jr}, \qquad 0 < x < r; \\ &= f_{r-1}(x-1|\theta_j)p_{jr}, \qquad x = r; \end{aligned} \tag{12}$$

where,

θ_j refer to a specific ability level x of examinees;
$f_r(x|\theta_j)$ is the distribution of correct scores over the first r items for examinees at the ability θ_j.

2.7 MODEL VALIDATION

There are three types of validity: criterion validity, content, and construct validity (PASQUALI; PRIMI, 2003; RICHARDSON, 2017).

According to Richardson (2017), 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 (NUNNALLY; BERNSTEIN, 1993).

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 (RICHARDSON, 2017).

Construct validity concerns the validation of a theory, which is reflected in a given instrument (RICHARDSON, 2017). It also can be defined as the extent to which a set of items, or tests, measures a latent trait (PASQUALI; PRIMI, 2003); 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 (NUNNALLY; BERNSTEIN, 1993).

Construct validity can be analyzed from several angles, from Classical Test Theory (CTT) and Item Response Theory (IRT) (PASQUALI; PRIMI, 2003). Urbina (2009) list as procedures that identify the latent trait, factor analysis, correlation with other tests, internal consistency, and convergent and discriminant validation. In our model, the latent trait is the support for awareness and collaboration provided by the collaborative environment.

2.7.1 Model reliability

The reliability of a set of items is one of the properties to evaluate the quality of the instrument. In classical test theory, reliability is generally defined as the ratio of the true score variance to observed variance or the squared correlation between true and observed scores, where the true score is the hypothetical average of the observed scores that would be obtained if the measurement were repeated over an infinite number of similar conditions (DEMARS, 2010).

IRT allows us to evaluate the assessment items' quality through θ, discrimination, and difficulty parameters. The test information function $I(\theta)$ is used to calculate standard error and reliability (DEMARS, 2010). Test information is a function of proficiency, and the items on the test vary with the proficiency level. The standard error of measurement is the inverse of the square root of information, so the greater the information, the smaller the standard error and the greater the reliability.

Similarly to classical approaches, one of the ways to check internal consistency is through Cronbach's alpha coefficient (α) (DEVELLIS, 2016). 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 (CRONBACH, 1951).

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 Equation 13. 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 (DEVELLIS, 2016).

$$\alpha = \frac{N\bar{c}}{\bar{v} + (N-1)\bar{c}} \tag{13}$$

where,

α 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 (θ)'s scale scores. We use the IRT technical properties of discrimination (a) and difficulty (b), combined with α coefficient, to assess the Awareness Assessment Model Instrument reliability and internal consistency.

2.7.2 Model dimensionality

An essential factor that corroborates the IRT model's validation is the latent trait's dimensionality, which, in our case, refers to the number of factors necessary to explain the variability of the data and constitute a hypothesis to be verified (SINGH, 2004). 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 (ANDRADE; TAVARES; VALLE, 2000). This factor is what is supposed to be measured by the assessment instrument. Schmitt (1996) 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 (SPENCER, 2004).

We used the Exploratory Factor Analysis and Confirmatory factor analysis to test dimensionality.

2.7.3 Exploratory Factor Analysis

Exploratory Factor Analysis (EFA) investigates the pattern of correlations between variables and uses these patterns of correlations to group the variables into factors. 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 common factors, called factors (HAIR et al., 2009). Thus, EFA evaluates the dimensionality of a series of items to identify the smallest number of latent traits that explain the correlation pattern (OSBORNE, 2014). The EFA model can be written according to Equation 14.

$$X_i = a_{i1}F_1 + a_{i2}F_2 + ...a_{im}F_m + e_i \quad (14)$$

with,

$$\mu = 0; \qquad \sigma = 1; \qquad i = 1, ..., p$$

where,

X_i is the $i - th$ standardized score;
p is the number of variables;
$a_{i1}, a_{i2}, ..., a_{im}$ are the factor loading for the i-th test;
$F_1, F_2, ..., F_m$ are the common uncorrelated factors;
e_i is an error specific to the $i - th$ test that is uncorrelated with any common factors.

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. 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 (PASQUALI; PRIMI, 2003).

Hair et al. (HAIR et al., 2009) suggest that the variables have communality greater than 0.5; however, in this research, we adopt commonality above 0.2 to not exclude essential assessment items. 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 (HAIR et al., 2009).

2.7.4 Confirmatory Factor Analysis

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 (THOMPSON, 2004; BROWN, 2015). Due to the sample size, we used the same data set for the EFA and CFA. In this configuration, Izquierdo, Olea, and Abad (2014) 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 (HAIR et al., 2009). In the Latent Root, the factors or components retained in the analysis with real data must have an eigenvalue higher than ones obtained randomly (LEDESMA; VALERO-MORA, 2019); 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 (HAIR et al., 2009).

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 (PASQUALI; PRIMI, 2003). 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 (PASQUALI; PRIMI, 2003). 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 (HAIR et al., 2009).

2.8 RELATED WORK

Some ways of evaluating awareness and collaboration were presented in the recent literature (MANTAU; BENITTI, 2022). 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 (MANTAU; BENITTI, 2022): by user experience (KIM et al., 2010), usability (BERKMAN; KARAHOCA; KARAHOCA, 2018), NASA-TLX user workload (GUTWIN; BATEMAN, et al., 2017), or ethnographic (HERSKOVIC; OCHOA; PINO; NEYEM, 2011) questionnaires; by ethnographic questionnaire combined with system logs and researcher's observations (MANTAU; BERKENBROCK; BERKENBROCK, 2014), and system logs (MANTAU; BERKENBROCK; BERKENBROCK, 2017); by participatory observations, nonstructured and mostly ad hoc interviews, and discussions (TALAEI-KHOEI; VICHITVANICHPHONG, et al., 2014); by semi-structured interview and a 7-point Likert scale questionnaire combined with statistical analysis (YANG, C.-L. et al., 2018); and by 7-points Likert scale ethnographic and usability questionnaire combined with researcher's observations, system logs, audio and video recordings (PROUZEAU; BEZERIANOS; CHAPUIS, 2018).

Frameworks, guidelines, design requirements, or groupware heuristics were also applied during development and evaluation, namely: a checklist to assess awareness support in collaborative systems (ANTUNES et al., 2014; ESPIRITO SANTO et al., 2018); set of requirements and assessment metrics (MANTAU; BERKENBROCK; BERKENBROCK, 2017); usability groupware heuristics for mobile environments (DE ARAÚJO et al., 2014); frameworks or taxonomies (GALLARDO; MOLINA, et al., 2011; SOUZA; BARBOSA, 2015; COLLAZOS et al., 2019; NIEMANTSVERDRIET et al., 2019); or questionnaires, laboratory testing, heuristic evaluation, automatic logging, and eye-tracking techniques (MOLINA et al., 2015).

Steinmacher, Gerosa, and Chaves (2010) and Steinmacher, Chaves, and Gerosa (2013) systematically reviews and maps the literature on awareness support in distributed software development; however, they do not study other collaborative systems. In general, most primary studies (79%) focus on introducing a new tool with some awareness support, and there is a lack of studies and tools giving solutions to awareness support.

Antunes et al. (2014) 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 Espirito Santo et al. (2018), where the authors investigate awareness support in the context of software engineering development.

Molina et al. (MOLINA et al., 2015) proposed an evaluating 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).

Collazos et al. (2019) carried out a systematic mapping study of frameworks and design processes of awareness support. This article reviewed which types of awareness information have been useful to the collaborative systems field and identified three types of context information sources: people, tasks or projects, and resources, such as workspace objects. The authors elaborated a descriptive theory for collaborative systems development to assist engineers in incorporating awareness mechanisms by focusing on the aspects to be considered when designing and implementing awareness mechanisms in collaborative tools.

Souza and Barbosa (2015) 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, Molina, et al. (2011) 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).

Lopez and Guerrero (2017) focused on a systematic review of collaborative tools that use ubiquitous mechanisms to provide awareness in the collaborative domain.

Niemantsverdriet et al. (2019) 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 Souza and Barbosa (2015), Collazos et al. (2019) and Niemantsverdriet et al. (2019) 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. (2014) described a checklist to inspect the quality of awareness support, this focuses mainly on the development stages and software designers. In addition, the validity and reliability of the proposed instrument were not checked.

The model proposed by Molina et al. (MOLINA et al., 2015) 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 (NIEMANTSVERDRIET et al., 2019), and it remains necessary to identify awareness evaluation criteria and establish quality indicators for collaborative environments (PROUZEAU; BEZERIANOS; CHAPUIS, 2018).

2.9 CONSIDERATIONS

This Chapter presents the most important theoretical pillars and topics related to this research. Firstly, we present the key concepts and definitions of the collaborative systems field and their intrinsic relationship with awareness. These awareness aspects will be presented in a detailed view in our systematic mapping study (see Chapter 3). In a second moment, we present the main approaches necessary to construct our assessment model, comprising the procedures necessary to create the model's conceptual view through an awareness taxonomy definition and techniques and mathematical models used to construct the assessment model.

Starting from constructing an awareness assessment model that considers the awareness support provided from the participant's perspective, we adopted the Item Response Theory as a statistical approach. As identified in the related works analysis, no awareness assessment approach was constructed in this perspective. A deeper understanding of the state of the art will be presented through our systematic mapping (Chapter 3).

We chose the IRT statistical approach due to the notable advantages of this approach compared to classical theory (TCT) like *i)* item statistics are independent (invariant) of the sample from which they were estimated; *ii)* the respondents' scores are independent of the difficulty of the test; *iii)* test analysis does not require rigorous parallel testing to assess reliability; and *iv)* item and respondents are positioned on the same scale.

In this work, we assume awareness as a participant's mental state (participant's consciousness). Thus, the assessment perspective is as important as the procedures and evaluation items; therefore, a proper awareness assessment requires considering the participant's skills and knowledge in assimilating awareness information/mechanisms provided, its context, expectations, and individual interaction/collaboration needs.

Adopting the IRT model becomes an interesting strategy, as it allows us to access the latent trail and build an ability scale from the participants' perspectives/skills. Literature confirms the effectiveness of the IRT in measuring latent traits (ANDRADE; TAVARES; VALLE, 2000) and allows the assessment of the phenomenon being studied with more accuracy and consistency and, therefore, provides relevant information and interpretations of the phenomenon under study (BORTOLOTTI et al., 2012).

This way, we assume IRT to be an appropriate statistical strategy for estimating the latent trait and constructing our awareness scale. Considering the degree of individualization of "what awareness is" and "how awareness is understood" by each participant – which could be difficult to represent in classical statistical models (TCT) –, we believe that the basis for properly assessing awareness support involves recognizing different abilities, understandings, or perspectives of those involved and incorporate them into the assessment.

As a result, our process seeks to establish an innovative approach to assessing awareness, wherein the participant's perspective/view becomes a fundamental part of the assessment process, thus contributing directly to the result obtained by the participant in the awareness support assessment.

3 SYSTEMATIC MAPPING STUDY

In science and engineering, systematically describing and organizing the investigated subjects help advance the field's knowledge (USMAN et al., 2017). Systematic mapping studies allow researchers to provide an overview of a research area by classification (KITCHENHAM; CHARTERS, 2007). Our study adopted the guidelines presented in Petticrew and Roberts (2006), Kitchenham and Charters (2007) and Petersen, Vakkalanka, and Kuzniarz (2015).

We performed a systematic mapping study to identify the state-of-the-art approaches adopted to develop and evaluate collaborative systems, address awareness and collaboration concepts (model 3C), and identify related challenges and limitations. This systematic mapping study aims to answer two main research questions:

Q1 *What are the approaches (e.g., models, methodologies, or processes) used in the development and evaluation of awareness and collaboration aspects in groupware systems?* (Section 3.3).
Specifically, we aim to identify:

- *How was considered the Cs of the 3C model? Does the model consider aspects of communication, coordination, and cooperation? How are they related?* (Section 3.3.1).
- *How was awareness related to the approach? What awareness types were reported? What elements were used to support it?* (Section 3.3.2).
- *Which time/space scenario was considered? How was the interaction performed (synchronous or asynchronous)? How were the participants geographically located?* (Section 3.3.3).
- *What were the research methods and techniques adopted? How was the approach evaluated or validated? What were the evaluation methods and instruments adopted?* (Section 3.3.4).

Q2 *What are awareness and collaboration issues and how do we understand them?* (Section 3.4).
We seek to identify:

- *Were non-functional or contextual aspects that impact system usage reported?* (Section 3.4.1).
- *What are the challenges and limitations of evaluating awareness and collaboration?* (Section 3.4.2).

The key terms follow the PICO structure (population, intervention, comparison, and outcomes), as presented by Petticrew and Roberts (2006). We applied the logical operator OR between key terms and the logical operator AND between the PICO dimensions, as exposed in Table 1. We performed the initial search using the IEEEXplore, ACM Digital Library, Scopus, Science Direct, Engineering Village, and Web of Science search engines, adjusting the query to each syntax.

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

3.1 EXECUTION PROTOCOL

We conducted the mapping study in a four-step path:

Step 1. Executing the query. In this step, the query was adjusted according to each search engine and executed by searching papers with correspondence in their title, abstract, or keywords. Then, we filtered the results obtained by the inclusion criteria (IC):

IC1. Works between 2010-2020;

IC2. Written in English or Portuguese;

IC3. Published in Journals or Proceedings.

Step 2. Applying the exclusion criteria. Based on the results obtained in the previous step, we applied the following exclusion criteria (EC):

EC1. Exclude 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. Exclude papers that did not address the groupware context, 3C model, or awareness;

EC5. Exclude papers that did not present a groupware development or evaluation model.

We considered papers addressing heuristic assessment methods, design guidelines, requirements or assessment approaches, or papers presenting awareness elements.

Step 3. Data extraction. Data from selected papers is extracted using the data extraction form (DEF). We use 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 Wohlin (2014). In the forward approach, we considered the Google Scholar indexation because it covers most of the databases used in the primary source. We performed the snowballing iteratively until we included no new paper, using both techniques in each iteration. Over the results, we applied the ICs of step 1 and executed the steps 2 and 3.

3.2 SYSTEMATIC MAPPING RESULTS

Executing the systematic mapping protocol obtained 1140 initial results. Fig. 8 shows the systematic mapping results, inspired in the PRISMA flowchart (HADDAWAY et al., 2022).

Performing steps 1 and 2 of the protocol, we selected 28 initial papers, as shown in Table 2. Using these 28 papers, we performed the backward and forward snowballing techniques, and the references and citations of each work were manually collected using the Google Scholar search engine. One by one, the records were manually transcribed to a corresponding spreadsheet, totaling 3180.

We performed three snowballing iterations until no new paper was found, as shown in Table 3. The column BSB represents the execution of the backward technique, and the column FSB represents the forward one. We analyzed 4320 records and selected 42 papers (see Table 4), considering 1140 results obtained in the search engines and 3180 results from snowballing techniques.

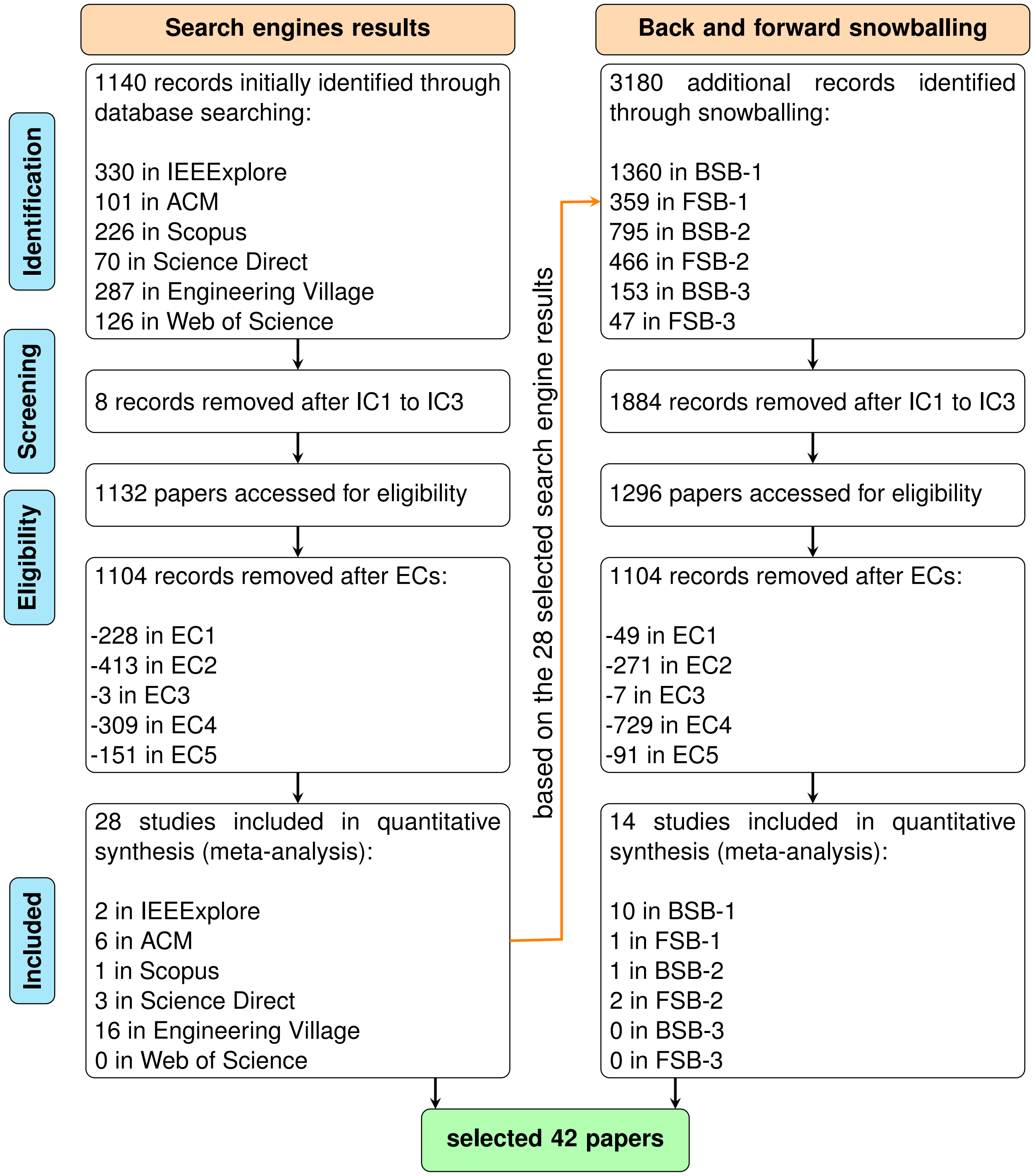


Figure 8 – Systematic mapping results (PRISMA flowchart)

Table 2 – Results obtained using search engines

	IEEE	ACM	Scopus	Science Direct	Eng. Village	Web of Science	total
initial	330	101	226	70	287	126	1140
IC filter	-0	-0	-4	-4	-0	-0	-8
EC1 filter	-82	-43	-10	-0	-85	-5	-228
EC2 filter	-52	-0	-167	-31	-83	-80	-413
EC3 filter	-1	-0	-2	-0	-0	-0	-3
EC4 filter	-149	-25	-32	-24	-44	-35	-309
EC5 filter	-44	-27	-7	-8	-59	-6	-151
total	2	6	1	3	16	0	28

Table 3 – Results obtained using snowballing

	BSB-1	FSB-1	BSB-2	FSB-2	BSB-3	FSB-3	total
initial	1360	359	795	466	153	47	3180
IC filter	-1004	-39	-681	-26	-134	-0	-1884
EC1 filter	-2	-34	-0	-13	-0	-0	-49
EC2 filter	-70	-63	-26	-85	-7	-20	-271
EC3 filter	-1	-3	-0	-3	-0	-0	-7
EC4 filter	-129	-213	-57	-294	-10	-26	-729
EC5 filter	-76	-9	-1	-5	-0	-0	-91
total	10	1	1	2	0	0	14

Table 4 – Selected papers

#	Year	Authors
1	2010	Idrus et al. (2010)
2	2010	Jonasson (2010)
3	2010	Kim et al. (2010)
4	2010	Omoronyia et al. (2010)
5	2010	Pelegrina et al. (2010)
6	2011	Gallardo, Molina, et al. (2011)
7	2011	Herskovic, Ochoa, Pino, and Neyem (2011)
8	2011	Hincapie-Ramos, Voida, and Mark (2011)
9	2011	Seebach, Beck, and Pahlke (2011)
10	2012	Altenburger et al. (2012)
11	2012	Antunes et al. (2012)
12	2012	Cruz et al. (2012)
13	2012	Neyem et al. (2012)
14	2012	Reinhardt et al. (2012)
15	2012	Rocker (2012)
16	2012	Talaei-Khoei, Ray, et al. (2012)
17	2012	Yuill and Rogers (2012)
18	2013	Belkadi et al. (2013)
19	2013	Decouchant et al. (2013)
20	2013	Gross (2013)
21	2013	Herskovic, Ochoa, Pino, Antunes, et al. (2013)
22	2013	Poulovassilis and Xhafa (2013)
23	2013	Xhafa et al. (2013)
24	2014	Antunes et al. (2014)
25	2014	De Araújo et al. (2014)
26	2014	Mantau, Berkenbrock, and Berkenbrock (2014)
27	2014	Talaei-Khoei, Vichitvanichphong, et al. (2014)
28	2015	Poulovassilis, Xhafa, and O'Hagan (2015)
29	2015	Souza and Barbosa (2015)
30	2015	Thomas, Botha, and Van Greunen (2015)
31	2016	Tripathi (2016)
32	2017	Fauzi, Sobri, and Suali (2017)
33	2017	Gutwin, Bateman, et al. (2017)
34	2017	Teruel et al. (2017)
35	2018	Berkman, Karahoca, and Karahoca (2018)
36	2018	Espirito Santo et al. (2018)
37	2018	Gallardo, Bravo, and Molina (2018)
38	2018	Mantau, Berkenbrock, and Berkenbrock (2017)
39	2018	Prouzeau, Bezerianos, and Chapuis (2018)
40	2018	Chi-Lan Yang et al. (2018)
41	2019	Collazos et al. (2019)
42	2019	Niemantsverdriet et al. (2019)

3.3 AWARENESS AND COLLABORATION APPROACHES

Q1 – What are the approaches (e.g., models, methodologies, or processes) used in developing and evaluating awareness and collaboration aspects in groupware systems?

The first main research question identifies the approaches (e.g., models, methodologies, or processes) used in developing and evaluating awareness and collaboration (model 3C) in groupware systems. To answer this question correctly, we factored it into four aspects: the 3C model, awareness, time-space, and research methods/techniques adopted.

3.3.1 3C Model

– How were considered the Cs of the 3C model? Does the model consider aspects of communication, coordination, and cooperation? How are they related?

The 3C dimensions were related in 24 papers (57%): coordination was the most reported dimension, being addressed in 20 papers (48%); communication was reported in 17 papers (40%); cooperation was covered in 9 works (22%). Figure 9 presents a detailed classification according to 3C model dimensions. This scenario demonstrates the high relationship between Cs and a few approaches addressed individually in each dimension (only ten works in this way). Furthermore, no studies considered communication or coordination and cooperation dimensions. Cooperation is a higher-level process guaranteed by communication and coordination (GUTWIN; BARJAWI; PINELLE, 2016).

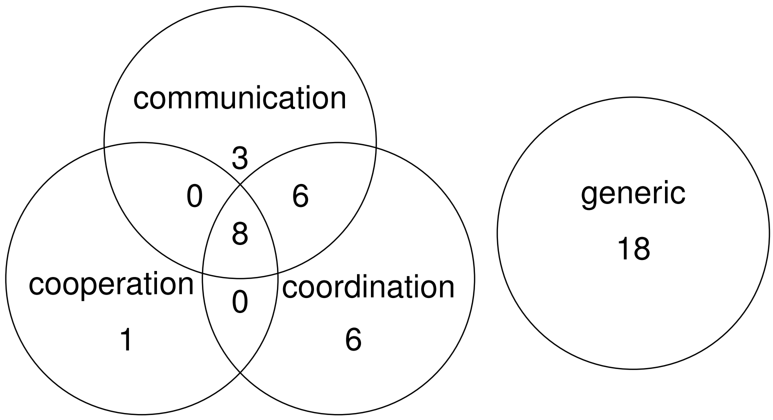


Figure 9 – Paper's classification according to 3C model

We observed the approaches broadly did not explicitly consider Cs of the 3C model. 18 of the selected papers (43%) presented a generic approach or did not specify exactly which communication, coordination, and cooperation aspects were considered. In these cases, the 3C model was implicitly related to the approach, mainly by providing awareness information/elements through the interface and providing basic aspects of communication, coordination, and cooperation.

3.3.2 Awareness

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

The results were manually analyzed, and related findings were extracted in a separate spreadsheet. Then, these were organized to establish standard terminology and remove the inconsistencies and redundancies. A terminology control and a relationship mapping between awareness views and their design elements were carefully established, providing a complete and unified view of different awareness classifications presented in the literature.

We used classifications presented by Gutwin and Greenberg (2002), Antunes et al. (2014) and Espirito Santo et al. (2018) as a starting point. To better understand awareness implications on the design, development, and evaluation of groupware, we have established a broad conceptual awareness framework consisting of five dimensions necessary in collaborative applications: contextual, collaboration, situational, workspace, and historical awareness.

The five awareness dimensions consist of 17 design categories and 92 supporting awareness elements or widgets. The awareness information related to each dimension was classified considering the 5W+1H framework (GUTWIN; GREENBERG, 2002). We presented a comprehensive but not exhaustive list of design categories and awareness elements necessary to design, develop, use, and evaluate groupware systems. Figures 59 to 14, in Appendix A contains a compiled view of these awareness structures.

Contextual Awareness view. It represents the notion of physical and virtual spaces, their topology, interaction ways, and mobility issues, as presented in Figure 10. In Appendix A Figures 59 and 60, we detail this awareness view, its design categories, and awareness elements.

This awareness view comprises four design categories and 22 design elements. It allows a group to maintain a sense of what is happening in virtual space and covers concepts of group navigation, physical/virtual spaces, spatiality, and mobility.

Navigation category represents information that assists participants through the shared environment. It contains six design elements: voice cues, portholes/ peepholes, eye-gaze cues, map views, viewports/teleports, and group and public objects. *Physicality/Virtuality* category represents information about the user's physical and virtual perspective. It contains six design elements: constraints, places, topology, attributes, location, and relationship. *Spatiality* category represents information that assists users in locating themselves in a shared environment. It contains seven design elements: distances, view, reach, orientation, territoriality, range of attention, and movement. *Mobility* 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: device mobility, user modality, and autonomy.

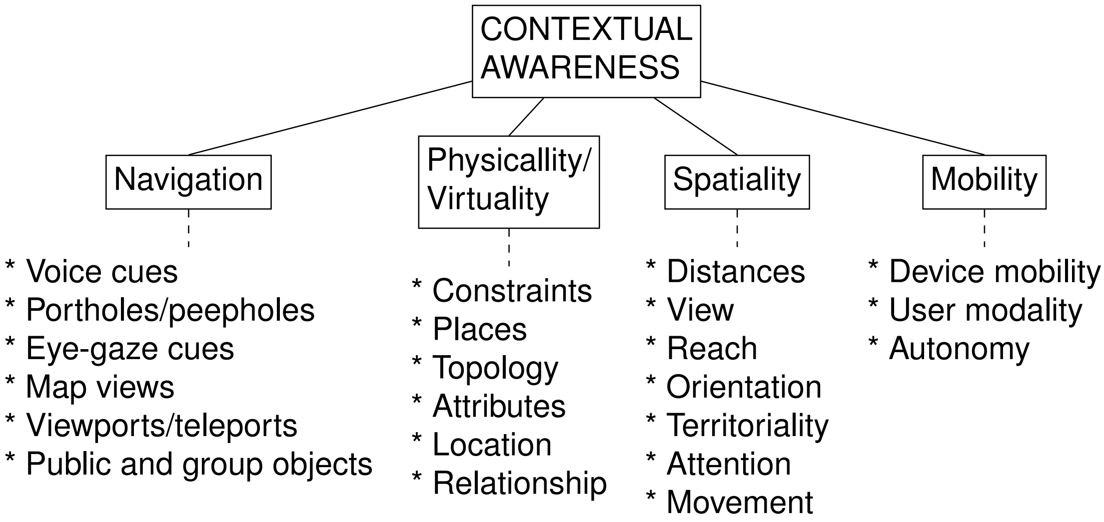


Figure 10 – Contextual awareness view

Collaboration Awareness view. It refers to members' perception of group availability, structure, and interaction. It consists of four design categories and contains 20 design elements. Collaboration awareness considers group identity, group capabilities, communication ways, and their status, as presented in Figure 11. In Appendix A Figures 61 and 62, we detail this awareness view, its design categories and awareness elements.

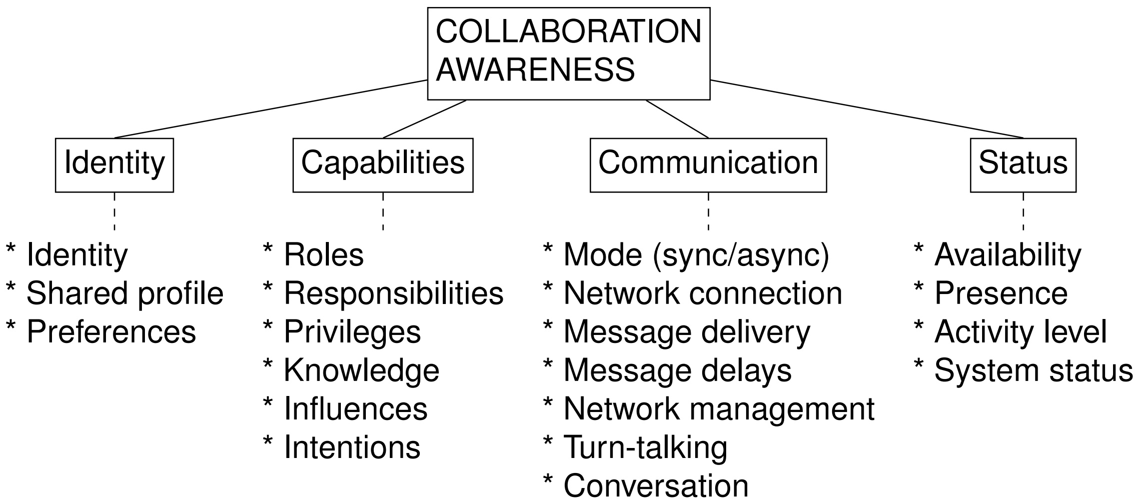


Figure 11 – Collaboration awareness view

Identity category composes the individual profile. It contains three design elements: identity, shared profile, and preferences. *Capabilities* category involves 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, responsibilities, privileges, individual/group knowledge, influences, and intentions. *Communication* 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), network connectivity, message delivery, message delays, network management, turn-talking, and conversation. *Status* category presents information allowing monitoring the current situation/availability of participants, system, task, and environment. It contains four design elements: availability, presence, activity level, and system status.

Situational Awareness view. It refers to the generalization of the notion of the workspace and concerns in understanding the environment, its elements, events, and actions performed. In short, it represents what is going on. This gives us understanding, sense-making, membership, and social information about group members. We organized this view into four design categories containing 21 elements, as presented in Figure 12. In Appendix A Figures 63 and 64, we detail this awareness view, its design categories, and awareness elements.

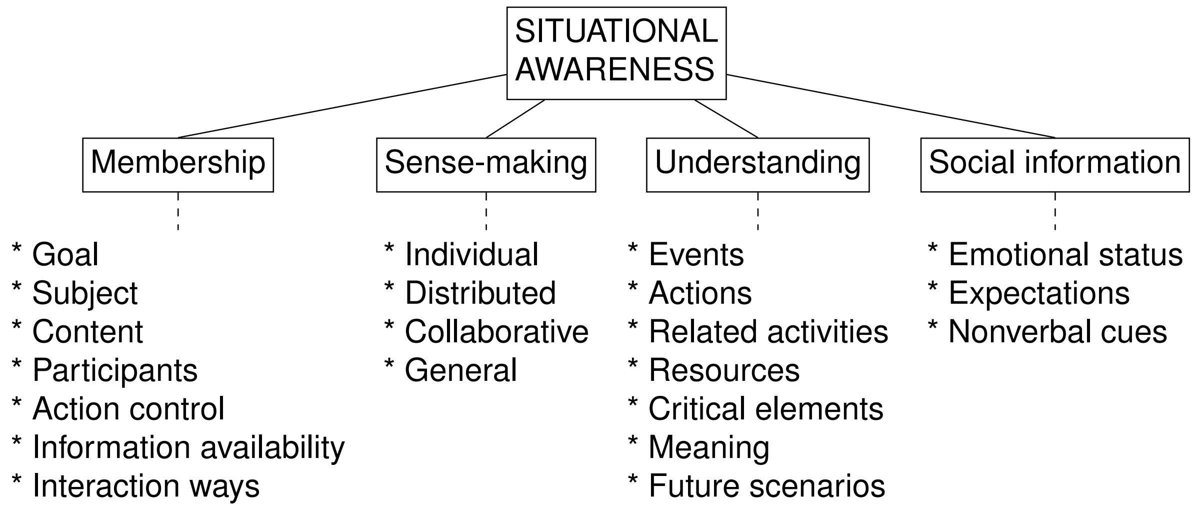


Figure 12 – Situational awareness view

Understanding category refers to a global perception of the environment. It contains seven design elements: events, actions, related activities, resources, critical elements, meaning, and future scenarios. *Sense-making* category provides insights into what happens and how individual, coordinated, and collaborative efforts influence group decision-making. It contains four design elements: individual sense-making, distributed sense-making, collaborative sense-making, and general sense-making. *Membership* category refers to the knowledge of belonging to or being a part of a group that allows constructing future scenarios and global knowledge of the environment. It contains seven design elements: goal, subject, content, participants, action control, information availability, and interaction ways. *Social information* category represents information about the participants' social perspective, emotions, expectations, premises, or other nonverbal cues that help to understand their actions, attitudes, or behavior. This information allows better design of future interactions between them and assists in identifying the difficulties and limitations on which you need assistance. It contains three design elements: emotional status, expectations, and other nonverbal cues.

Workspace Awareness view. It 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 three design categories and 18 design elements, representing tasks, interaction, and activities interdependence, as presented in Figure 13. In Appendix A Figures 65 and 66, we detail this awareness view, its design categories, and awareness elements.

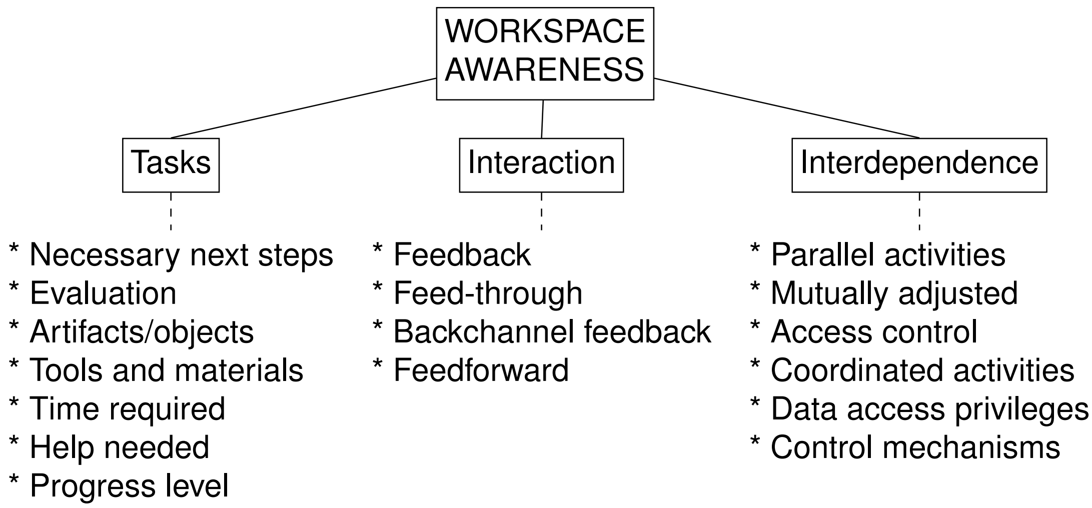


Figure 13 – Workspace awareness view

Tasks refers to activities, tasks, objects, and other elements existing in a shared environment. It contains eight design elements: necessary next steps, evaluation, tools and materials, time required, tasks completed, help needed, artifacts and objects, and progress level. *Interaction* represents the responses from individuals, others, or group actions through a groupware system that allows users to understand the effects of interaction. It contains four design elements: feedback, feed-through, backchannel feedback, and feedforward. *Interdependence* represents the relationship and dependency between activities, tasks, or shared objects, rules, precedence, or constraints imposed to their realization. It contains six design elements: parallel activities, mutually adjusted access control, coordinated activities, data access privileges, and control mechanisms.

Historical Awareness view. We consider historical awareness as a temporal perspective, representing changes made in the context, workspace, or group structure during the collaboration process. This group comprises two design categories and 11 design elements, as shown in Figure 14. It consists of a historical view of actions performed by users individually or by group and workspace changes over time. In Appendix A Figure 67, we detail this awareness view, its design categories, and awareness elements.

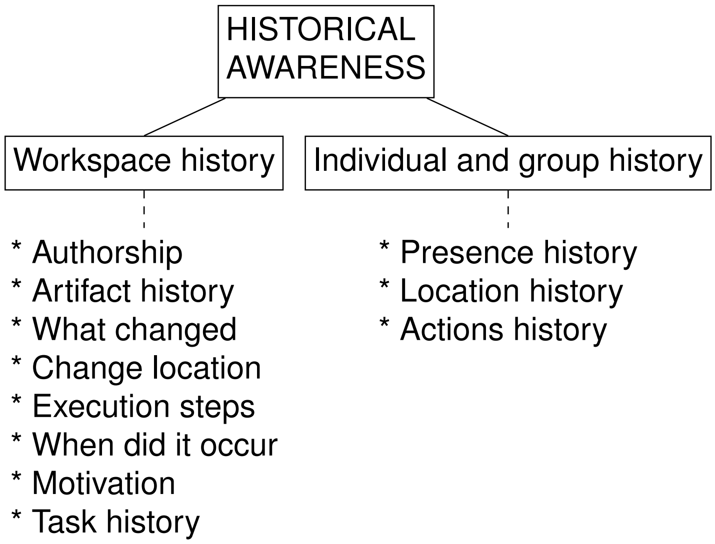


Figure 14 – Historical awareness view

Workspace history category gives us a temporal view of actions taken on objects belonging to the shared workspace. It contains eight design elements: authorship, artifact history, what changed, change location, execution steps when it occurred, motivation, and task history. *Individual and group history* category provides a temporal view of the individual and group cooperation process. It contains 3 design elements: presence, location, and actions.

3.3.3 Time/Space

– Which time/space scenario was considered? How was the interaction performed (synchronous or asynchronous)? How were the participants geographically located?

Regarding the space/time classification defined by Bullen and Johansen (1988) and Ellis, Gibbs, and Rein (1991), we found that 22 papers (52%) were built in the context of synchronous distributed interaction environments; 17 papers (40%) did not consider the time/space, presenting a generic approach; and just three papers (7%) were classified in other space/time quadrants.

This concern with synchronous distributed interaction remains nothing new. Distributed environments present some challenges in awareness support. When people move from face-to-face to a distributed setting, many things do change that impair people's abilities to maintain awareness (GUTWIN; GREENBERG; ROSEMAN, 1996). Compared to face-to-face interaction, the provided awareness cues are relatively incomplete, uncertain, and poor (OULASVIRTA, 2008). The perceivable environment means communication and perceptual and physical abilities over shared artifacts shrink drastically. Providing similar face-to-face interaction mechanisms in a distributed environment is a well-known but not fully solved problem. Face-to-face interaction is hard to represent since it can be subtle, hard to observe, or buried within several layers of inference (GUTWIN; GREENBERG; ROSEMAN, 1996).

3.3.4 Research methods and techniques

– What were the research methods/techniques adopted? How was the approach evaluated or validated? What were the evaluation methods and instruments adopted?

We analyzed the papers regarding their type (WIERINGA et al., 2006), research methods (PETERSEN; VAKKALANKA; KUZNIARZ, 2015), instruments, materials, and data collection techniques adopted. Figure 15 summarizes the relationship between the research method used (X-axis) and paper classification (Y-axis).

Research methods. The main strategy adopted in the research approach is the development of collaborative applications and evaluation involving users (22 papers), either by experiments (2 papers) or case studies (20 papers). Nine papers presented guidelines through a literature review; 4 papers carried out surveys through questionnaires in specific groups and using a qualitative or specialist analysis; and seven papers adopted other research approaches, such as technical review, comparative study, or heuristic evaluation.

Papers classification. Our analysis identified 28 research papers, four solution proposal papers, two experience papers, one opinion paper, and none addressed validation research.

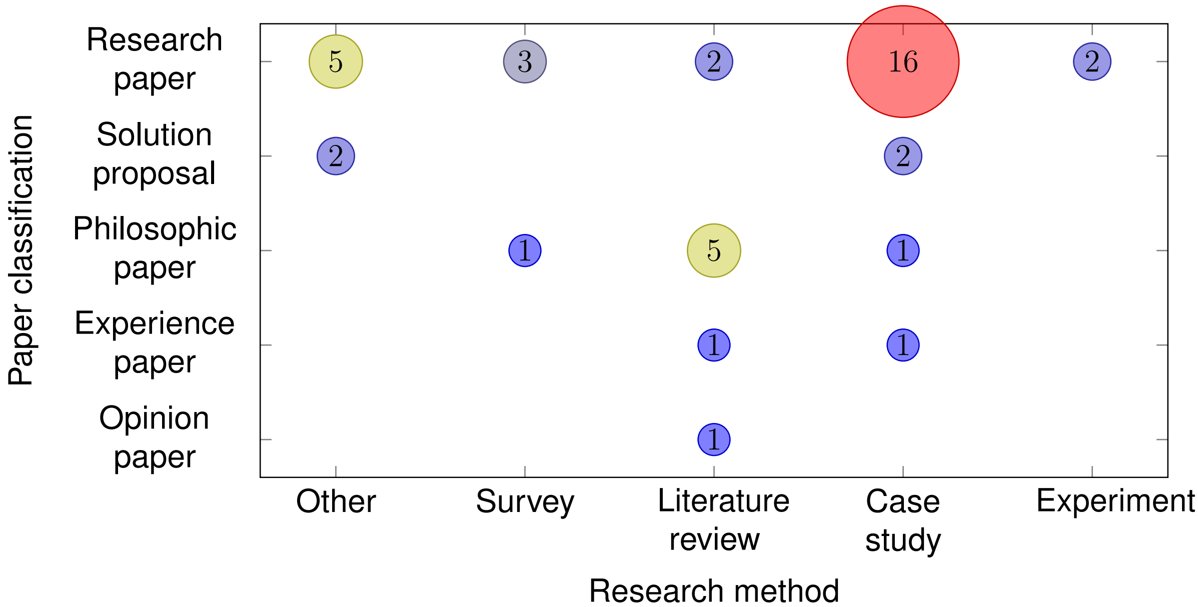


Figure 15 – Papers classification and research methods

Instruments and materials. Data were collected using questionnaires, interviews, brainstorming, focus groups, conceptual modeling, direct observation, system logs, and static/dynamic analysis of a system.

Questionnaires were the main data collection tool reported. We identified the following approaches: by using user experience (KIM et al., 2010), usability (BERKMAN; KARAHOCA; KARAHOCA, 2018), NASA-TLX user workload (GUTWIN; BATEMAN, et al., 2017), or ethnographic (HERSKOVIC; OCHOA; PINO; NEYEM, 2011) questionnaires; by using ethnographic questionnaire combined with system logs and researcher's observations (MANTAU; BERKENBROCK; BERKENBROCK, 2014), and system logs (MANTAU; BERKENBROCK; BERKENBROCK, 2017); by using participatory observations, non-structured and mostly ad-hoc interviews, and discussions (TALAEI-KHOEI; VICHITVANICHPHONG, et al., 2014); by using semi-structured interview and a 7-points Likert scale questionnaire combined with statistical analysis (YANG, C.-L. et al., 2018); and by using 7-points Likert scale ethnographic and usability questionnaire combined with researcher's observations, system logs, and audio and video recordings (PROUZEAU; BEZERIANOS; CHAPUIS, 2018).

Frameworks, guidelines, design requirements, or groupware heuristics were used during development and evaluation, namely: checklist to assess awareness support in groupware systems (ANTUNES et al., 2014; COLLAZOS et al., 2019); set of requirements and assessment metrics (MANTAU; BERKENBROCK; BERKENBROCK, 2017); usability groupware heuristics for mobile environments (DE ARAÚJO et al., 2014); and frameworks or taxonomies (SOUZA; BARBOSA, 2015; GALLARDO; MOLINA, et al., 2011; COLLAZOS et al., 2019; NIEMANTSVERDRIET et al., 2019).

3.4 ISSUES AND CHALLENGES

Q2 – What are the awareness and collaboration issues, and how do we understand them?

The second main research question involves identifying awareness and collaboration issues and how to grasp them. According to Schmidt and Randall (2016), since the origin of the CSCW field, there is no consensus about it. We do not try answering this question directly here. However, we seek to present a compiled set of aspects and problems reported in the selected papers that provide insights into possible answers. First, we identified various issues or non-functional constraints supporting awareness and collaboration that impact collaborative system use. These were grouped into awareness, collaboration, interaction, technical, and design issues. Second, we identified challenges and limitations in evaluating awareness and collaboration aspects.

3.4.1 Non-functional constraints

– Were non-functional or contextual aspects of system usage reported?

We have identified several issues related to support for awareness and collaboration aspects. These findings were compiled into six categories: awareness, collaboration, interaction, technical, contextual, and design issuers.

3.4.1.1 Awareness Issues

Awareness support considers awareness obtainment, creation, modeling, labeling, distribution, and manipulation (TALAEI-KHOEI; RAY, et al., 2012; POULOVASSILIS; XHAFA, 2013). There is a need to represent awareness information to the participants, demanding to consider the types of awareness information needed (NIEMANTSVERDRIET et al., 2019), possible mechanisms for its representation (YANG, C.-L. et al., 2018), and different personas (participants roles and rules) to model shared representations (BELKADI et al., 2013).

The light-weight mechanisms and filtering techniques allow us to exchange valuable and up-to-date information, reducing the overhead of notifications to the user and intrusiveness (POULOVASSILIS; XHAFA, 2013; POULOVASSILIS; XHAFA; O'HAGAN, 2015; MANTAU; BERKENBROCK; BERKENBROCK, 2017; NIEMANTSVERDRIET et al., 2019). Restrictions such as information overload (MANTAU; BERKENBROCK; BERKENBROCK, 2014), effort load required (NIEMANTSVERDRIET et al., 2019), and cognitive load must be transposed (MANTAU; BERKENBROCK; BERKENBROCK, 2017).

3.4.1.2 Collaboration issues

Collaboration involves communication, coordination, and cooperation processes. In collaboration process, participants need different interaction modalities (ROCKER, 2012; SOUZA; BARBOSA, 2015; NIEMANTSVERDRIET et al., 2019) and suitable communication mechanisms (CRUZ et al., 2012; DE ARAÚJO et al., 2014; MANTAU; BERKENBROCK; BERKENBROCK, 2017).

The communication process involves implicit (non-verbal), explicit (text, messages, annotations), and secondary or alternative notations (COLLAZOS et al., 2019). The coordination process involves group management (XHAFA et al., 2013; DE ARAÚJO et al., 2014; MANTAU; BERKENBROCK; BERKENBROCK, 2017) and the use of mechanisms to enable participants to organize themselves into a shared environment (CRUZ et al., 2012). Time, tasks, shared artifacts' management, coordination protocols, and control models are additionally required (CRUZ et al., 2012).

The cooperation among participants involves considering group issues like its composition, structure (leadership and hierarchy), and autonomy (CRUZ et al., 2012;

TERUEL et al., 2017), participants' behavior, their background/grounding and attitude towards technology (CRUZ et al., 2012; BERKMAN; KARAHOCA; KARAHOCA, 2018) and group consistency and standards (GROSS, 2013; DE ARAÚJO et al., 2014; MANTAU; BERKENBROCK; BERKENBROCK, 2017).

3.4.1.3 Interaction issues

Cooperative work involves interaction among participants. This process is complex, and it is necessary to consider some tasks, sociotechnical, and outcome concerns. Performing collaborative tasks involves participants' creativity, intellect, commitment, and decision-making problem-solving competencies (CRUZ et al., 2012; MANTAU; BERKENBROCK; BERKENBROCK, 2017), and the progressive self-evaluation (SOUZA; BARBOSA, 2015). The tasks' viscosity, visibility, complexity, goals, and physical setting should be considered (SOUZA; BARBOSA, 2015; TERUEL et al., 2017).

People are different in their social norms, previous experience (CRUZ et al., 2012) understandings (NIEMANTSVERDRIET et al., 2019), levels of familiarity (YUILL; ROGERS, 2012), and attention level (GUTWIN; BATEMAN, et al., 2017). They have different capabilities in representing, understanding, and projecting human actions through interface (MANTAU; BERKENBROCK; BERKENBROCK, 2017). Some sociotechnical factors influence the interaction: the participants' motivation and mediation, group vitality, patterns, scripts and techniques, division of labor, and sharing of participants' viewpoints, knowledge, work, tasks, and goals (CRUZ et al., 2012; MANTAU; BERKENBROCK; BERKENBROCK, 2017).

The outcomes are related to the quality of group performance, expectations and satisfaction with system use, appreciation of group membership, individual breakdowns, individual rewards, organizational results, learning monitoring, interaction degree, and group integration (CRUZ et al., 2012; BERKMAN; KARAHOCA; KARAHOCA, 2018).

3.4.1.4 Technical issues

Hardware and software can help to change the constraints on mechanisms of behavior identified (YUILL; ROGERS, 2012). In a generic view of these technical constraints, it is necessary to incorporate design considerations to different types of applications and contexts (NIEMANTSVERDRIET et al., 2019) to deal with device and environmental restrictions such as screen dimensions, power limitation, network data use, or communication delays (MANTAU; BERKENBROCK; BERKENBROCK, 2017), to support long periods of disconnection in synchronous environments (GUTWIN; BATEMAN, et al., 2017), to provide user mobility, and to avoid system intrusiveness during user actions (MANTAU; BERKENBROCK; BERKENBROCK, 2017).

At the application layer, it is necessary to consider the architecture, functional and quality properties, group processes support, core functionality, supported actions, proper alert mechanisms, intelligent or semi-intelligent software components, actors tools, and roles involved (CRUZ et al., 2012).

At the network layer, there are concerns like automatic connection service and device discovery (HERSKOVIC; OCHOA; PINO; NEYEM, 2011; NEYEM et al., 2012), network limitations (POULOVASSILIS; XHAFA, 2013; MANTAU; BERKENBROCK; BERKENBROCK, 2017), message exchange (synchronous, asynchronous and pushing) (HERSKOVIC; OCHOA; PINO; NEYEM, 2011; NEYEM et al., 2012; CRUZ et al., 2012), dynamic network conditions and multiple nodes for sending information (passive and active mode) (POULOVASSILIS; XHAFA, 2013; POULOVASSILIS; XHAFA; O'HAGAN, 2015), data synchronization (XHAFA et al., 2013), and response time (MANTAU; BERKENBROCK; BERKENBROCK, 2017).

The mobile and tabletop platforms present challenges that must be overcome during the design and construction of groupware systems. In the mobile context, we have mobility, mixed-presence, simultaneous interaction, weak/tightly coupled interaction, and territoriality (KIM et al., 2010; MANTAU; BERKENBROCK; BERKENBROCK, 2017). In tabletop environments, it is necessary to support interpersonal interaction, fluid transitions between activities (personal and group work), use of physical objects, and transitions between tabletop collaboration and external work. Additionally, it is necessary to provide shared access to physical and digital objects, to establish appropriate arrangement of users, and to support simultaneous user actions (JONASSON, 2010).

3.4.1.5 Contextual issues

It consists of information gathering, representation, and dissemination (TALAEI-KHOEI; RAY, et al., 2012; ROCKER, 2012; MANTAU; BERKENBROCK; BERKENBROCK, 2017); and the peripheral information presentation (ROCKER, 2012). It is hard to represent the contextual/situational factors (rewards, budget, and training), cultural contexts (trust or equity), and group environment (competition, uncertainty, time pressure, and evaluative tone) (CRUZ et al., 2012).

Data consistency and availability involve data replication, caching, and conflict resolution strategies (HERSKOVIC; OCHOA; PINO; NEYEM, 2011; NEYEM et al., 2012). It is necessary to avoid the overload of irrelevant or loosely relevant information and deal with information uncertainty (COLLAZOS et al., 2019). Multi-version data management, fine-grain locking, user-level locking, application-defined transactions, operation semantics-based concurrency control, and notification messages are reported techniques that allow maintaining the level of consistency and availability of the shared data (ROCKER, 2012; MANTAU; BERKENBROCK; BERKENBROCK, 2014, 2017).

3.4.1.6 Design issues

It comprises general issues like usability constraints, user protection, information availability, and user mobility and flexibility. The usability constraints are related to the aspects of the visibility of system status, aesthetic and minimalist design, ease of input, viewing and screen reading, aesthetic, social and private conventions, tailoring, recognition, attention level, and transition among activities (DE ARAÚJO et al., 2014; SOUZA; BARBOSA, 2015; MANTAU; BERKENBROCK; BERKENBROCK, 2017).

Users' protection is related to the error management (DE ARAÚJO et al., 2014; SOUZA; BARBOSA, 2015; MANTAU; BERKENBROCK; BERKENBROCK, 2017; NIEMANTSVERDRIET et al., 2019), avoid intrusiveness (MANTAU; BERKENBROCK; BERKENBROCK, 2017), user control (ROCKER, 2012; NIEMANTSVERDRIET et al., 2019), user security and privacy (HERSKOVIC; OCHOA; PINO; NEYEM, 2011; NEYEM et al., 2012; MANTAU; BERKENBROCK; BERKENBROCK, 2017; NIEMANTSVERDRIET et al., 2019), authorization and access rules (KIM et al., 2010; CRUZ et al., 2012; POULOVASSILIS; XHAFA, 2013; PROUZEAU; BEZERIANOS; CHAPUIS, 2018), rules management (constraints, norms or work rules) (KIM et al., 2010; CRUZ et al., 2012; TERUEL et al., 2017), rules-based collaboration (KIM et al., 2010), context-dependent privacy profiles (ROCKER, 2012), and ad-hoc work session (HERSKOVIC; OCHOA; PINO; NEYEM, 2011).

Information availability refers to the data availability, their abstraction (sensor data, natural, multimedia) (HINCAPIE-RAMOS; VOIDA; MARK, 2011; SOUZA; BARBOSA, 2015; NIEMANTSVERDRIET et al., 2019), presentation (continuous, discrete or literal), delivery (always on, almost always on, on request or implicit interaction), symmetry (symmetric traceable, symmetric blind, asymmetric traceable or asymmetric blind), obtrusiveness (focal, selective focal, secondary appliance or peripheral) and temporal gradient (current, recent, historical or predicted availability) (HINCAPIE-RAMOS; VOIDA; MARK, 2011).

User mobility requires considerations about device heterogeneity (HERSKOVIC; OCHOA; PINO; NEYEM, 2011; NEYEM et al., 2012; MANTAU; BERKENBROCK; BERKENBROCK, 2017), operational and technological interoperability (XHAFA et al., 2013), and efficiency in resources usage (MANTAU; BERKENBROCK; BERKENBROCK, 2014). The flexibility includes automatic user detection, connection, and disconnection support (HERSKOVIC; OCHOA; PINO; NEYEM, 2011; MANTAU; BERKENBROCK; BERKENBROCK, 2017; NIEMANTSVERDRIET et al., 2019).

3.4.2 Evaluation challenges

–What are the challenges and limitations related to evaluating awareness and collaboration aspects?

Evaluating collaborative systems is more complex and challenging than conventional ones. We have compiled the four main reasons supporting this scenario here.

First, providing awareness and 3C model aspects while dealing with issues and challenges, as previously presented, represents a grand challenge in building groupware systems. In this context, there is a need to balance two main trade-offs: informativeness versus privacy: if current status of a person is visible enough to be helpful to others, it often violates that person's privacy (ROCKER, 2012); information versus overloading: the lack of awareness information may compromise group's activities, on the other hand, it is essential to avoid information overload, presenting just relevant information to user (MANTAU; BERKENBROCK; BERKENBROCK, 2014).

Second, due to the groupware evaluation being in more than one temporal dimension, it is complex to obtain data about each view in just one way (ANTUNES et al., 2012). Information about an individual is gathered focusing on events occurring in a time frame of a few minutes or even seconds; the group information is gathered addressing activities occurring in the range of several minutes and hours; and information regarding organizational impact concerns much longer time frames, usually in the order of days, months, and even years.

Third, research still fails to address conceptual frameworks covering the four trends: theoretical frameworks, context modeling, collaborative design, and awareness (BELKADI et al., 2013). 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 (CRUZ et al., 2012). 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 (CRUZ et al., 2012). It is hard to generate adaptation rules automatically, and no frameworks help designers to incorporate semi-automatically users' feedback (ALTENBURGER et al., 2012). According to Bravo et al. (2023), this problem remains true; it is necessary to fill the gap in methods and tools to guide and facilitate the design and development of the awareness support and build suitable awareness support for a groupware system according to the users' requirements and tasks.

Fourth, and most worrying, is the fact that few works present methods or processes that assist in providing aspects of awareness in groupware systems (COLLAZOS et al., 2019) and there are no standardized tests for awareness assessment (PROUZEAU; BEZERIANOS; CHAPUIS, 2018). There is a need to establish measures to assess awareness (PROUZEAU; BEZERIANOS; CHAPUIS, 2018), identify the criteria for achieving awareness, and establish indicators (NIEMANTSVERDRIET et al., 2019). Future research in this direction is necessary and will bring great advances towards designing, developing, and evaluating groupware systems.

3.5 PAPERS QUALITY ASSESSMENT

To increase confidence in our systematic mapping study, we rigorously assess the quality of the primary studies included. Many quality assessment checklists are established in the literature, each identifying a slightly different set of questions (KITCHENHAM; CHARTERS, 2007). Furthermore, there are no agreed-upon universal quality assessment checklists and procedures (SÁNCHEZ-GORDÓN; COLOMO-PALACIOS, 2019). We adopted the criteria presented by Dybå and Dingsøyr (2008a,b) for the papers' quality evaluation. The former is a well-known and widely used quality assessment checklist described in the literature, serving as a basis for recent guidelines for conducting systematic mapping studies like Kitchenham and Brereton (2013) and Petersen, Vakkalanka, and Kuzniarz (2015). Due to its adaptability, it has been the most prevailing quality checklist for systematic literature reviews in the software engineering field (YANG, L. et al., 2020). The latter defines four essential aspects that should be considered when assessing the quality of primary studies: reporting, rigor, credibility, and relevance. Table 5 contains the quality assessment questions.

Table 5 – Papers quality assessment questions

#	Description
Q1	*Study type*. Is it research work lessons learned or expert opinions?
Q2	*Objectives*. Were the research objectives, motivation, and results clearly defined?
Q3	*Context*. Was the context clearly described?
Q4	*Approach*. Is the research approach adequate to the research objectives? Were the adopted procedures presented and justified?
Q5	*Sample*. Is the sample adequate for the research objectives? Were the selection criteria described? Is the most appropriate approach for the proposed objectives?
Q6	*Control group*. Is there a control group for comparison? If so, what were the selection criteria? How representative are they to the population?
Q7	*Data collection*. Do the data support research questions? Were the measures taken? How was the data collected?
Q8	*Data analysis*. Was the data analysis conducted and described in detail? Are the data sufficient to support the reported findings? How were the results obtained verified?

continues on the next page

Table 5: Papers quality assessment questions (continuation)

Q9	*Impartiality/Neutrality*. What is the researcher's previous relationship with the context studied? Is this relationship adequate for the proposed study? Were the researcher's potential biases or influences analyzed during the study (formulation of research questions, selection, collection, and data analysis)?
Q10	*Findings*. Were the findings presented with evidence? Were the study limitations discussed? Were the conclusions reached by the work justified by the results obtained?
Q11	*Contributions*. Is there a discussion of the contributions? Are they relevant? Have current gaps or areas in which further research is required? Was it discussed how to transfer the results to other populations or contexts?

We assessed each primary study for quality simultaneously with the data extraction process. Each question was independently answered and assigned values 1 (answer or present), 0.5 (partially answer or presented), and 0 (does not answer/present). The final quality represents the sum of values attributed to each question. The title, abstract, and keywords were assessed from Q1 to Q3. The full text was assessed from Q4 to Q11. We grouped the results by papers' quality level considering the following quality scale: high (≥ 8); moderate (≥ 7 to < 8); low (≥ 6 to < 7); and very low or poor (< 6). From a general point of view of quality assessment, we classify nine papers as high quality, ten as moderate quality, 16 as low quality, and seven as very low or poor quality.

Reporting (criteria Q1-Q3). It represents the reporting quality, encompassing its reasons, aims, and context. This issue evaluation was very positive, satisfying all the criteria from Q1 to Q3. We believe it is because these aspects were handled as exclusion criteria during the protocol execution, respectively: type of study (EC1), objectives (EC5), and context (EC4).

Rigor (criteria Q4-Q8). It represents the rigor of research methods employed to establish the validity of data collection and analysis methods, focusing on whether the research method is thorough and appropriate. Generally, the studies were well evaluated concerning the research approach adopted (Q4). Few studies have presented the procedures for obtaining the sample utilized in the study (Q5) and data collection (Q7). These questions obtained low scores, indicating the need for greater efforts to present this information in future works. Control groups (Q6) were unreported in studies, partly due to it being inapplicable considering the research approach or method adopted. Regarding the data analysis (Q8), six studies (14%) did not perform or detail the procedures adopted, and fourteen studies (33%) presented a poor data analysis or were not detailed, making it difficult to analyze the processes involved in achieving these results.

Our findings corroborate to the Steinmacher, Chaves, and Gerosa (2013). They identified that 47 papers (52%) did not present experimental analysis, and 68 articles (75%) presented a new tool. Lopez and Guerrero (2017) identified that almost half of 83 papers (42%) did not specify the evaluation approach used to assess the system, and just 6% presented an example proposal. In our quality assessment, we identified 45% of papers with a substantial problem in data analysis and 9% just presenting a solution proposal paper, demonstrating the need for greater rigor in adopting methods, procedures, and materials when conducting studies.

Credibility (criteria Q9-Q10). It represents the credibility of study methods for ensuring the findings were valid, meaningful, and well-presented. The analyzed papers did not present the researcher's impartiality/neutrality (Q9). Although studies were well evaluated concerning the findings (Q10), in many of these, data were insufficient to support the findings due to a smaller sample of users and poor data analysis (Q8).

Relevance (criteria Q11). It represents the study's relevance to the software industry and research community. The contributions (Q11) presented by papers are interesting; however, a better rigor would more properly support it.

3.6 THREATS TO VALIDITY

Some procedures were adopted during the conduction of the research to mitigate threats to validity. These threats could compromise the quality and reliability of research results. We carried out our systematic mapping following an execution protocol documented and reviewed by a specialist. We followed the guidelines presented in Petticrew and Roberts (2006), Kitchenham and Charters (2007) and Petersen, Vakkalanka, and Kuzniarz (2015), including all necessary research steps. During execution, the researcher rigorously led all procedures defined in the protocol.

The search engines were carefully selected to ensure a broad range of relevant publications in the area. We considered six widely known search databases in software engineering (KITCHENHAM; CHARTERS, 2007). Besides, we used a non-restrictive search query combined with backward and forward snowballing techniques according to Wohlin (2014). It mitigates the risk that the primary studies obtained might not correspond to the current state of the art.

All steps performed and the results generated during the extraction and selection steps were documented, allowing the review of the procedures. We objectively defined the selection criteria (ICs and ECs). If there are doubts about applying the criteria, the paper was temporarily included in the result and assigned to be reviewed by another specialist.

A unique researcher's data extraction and analysis procedures may result in biased interpretation. In response, we meticulously followed the extraction procedures defined in the protocol and documented all stages of data collection. Seeking to minimize possible biases, we adopted a two-stage data extraction: Initially, we extracted information related to research questions in each selected work. Then, we synthesized the information obtained and reviewed the original papers.

3.7 DISCUSSION

The concepts of awareness and collaboration have been intrinsically related since the conception of the CSCW field, and its understanding has expanded in the same way as research in the field evolves. During the previous decades, we observed the scientific community's efforts toward establishing common sense about what awareness is, what it represents, and what it is related to. However, achieving an accurate and clear-cut definition of awareness is still very difficult or even impossible, as already identified by Gross (GROSS, 2013).

Understanding and providing aspects of collaboration also involves comprehensive knowledge of elements of awareness that support it. We do not envision ways to provide efficient aspects of communication, coordination, or cooperation without awareness being supported by a groupware system. Furthermore, provisioning adequate awareness mechanisms ensures the support for each C of the 3C collaboration model – it consolidates awareness as a cornerstone of collaborative environments.

In this Chapter, we conducted a systematic mapping study to identify approaches (models, methodologies, or processes) adopted in the design, development, and evaluation of groupware systems, addressing the awareness and collaboration (3C model) concepts in a meticulous investigation of the last ten years CSCW publications. We researched, collected, and analyzed 4320 articles using a detailed protocol, considering 42 identified publications. These works were the basis for answering two central awareness research questions.

In the first research question, we focused on awareness and collaboration approaches used in developing and evaluating collaborative systems, reporting considerations about the dimensions of the 3C Model involved, the relationship with each other, and awareness. In the second one, we identified a comprehensive set of non-functional or contextual aspects (context-aware) and challenges and limitations related to the awareness and collaboration evaluation.

3.7.1 Contributions

In this Chapter, we made two contributions toward a broader awareness and understanding. First, regarding awareness support, we identified a set of 92 supporting awareness elements or widgets organized into 17 design categories that address five main awareness dimensions. It represents a set of awareness elements synthesized concerning the consolidated classification of the 5W+1H framework (GUTWIN; GREEN-

BERG, 2002). Formerly, we focus on assisting in designing and evaluating collaborative applications by elaborating a conceptual framework of awareness maintenance.

Second, we gathered various awareness design issues, categorizing them into awareness, collaboration, interaction, technical, and design issues. It represents the fundamental non-functional constraints related to providing awareness elements through groupware interfaces, and it demonstrates the proper dimension of awareness support in collaborative environments. Some challenges in awareness and collaboration evaluation were provided, demonstrating that although the awareness problem is already well known in the literature, it remains without a universal answer – or even if it is possible to achieve.

In short, understanding the concepts presented here provides insights into two key objectives of this thesis. In the first place, we are working on consolidating an awareness taxonomy that encompasses our findings and unifies taxonomies already existent in literature (See Chapter 4). Second, we seek to establish an awareness assessment method or process to provide awareness aspects in collaborative systems.

3.7.2 Related publications

The results presented in this chapter were published in:

- MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. Awareness Support in Collaborative System: Reviewing last 10 years of CSCW Research. In: IEEE. 2022 IEEE 25th International Conference on Computer-Supported Cooperative Work in Design (CSCWD). IEEE, 2022. P. 564–569. DOI: 10.1109/cscwd54268.2022.9776091.

Part II

Model Construction

4 AWARENESS TAXONOMY

Our taxonomy definition method is based on Bailey's conceptual approach Bailey (1994) and combines the guidelines presented by Nickerson, Varshney, and Muntermann (2013), Usman et al. (2017) and Szopinski, Schoormann, and Kundisch (2019).

Bailey's conceptual method uses the deductive approach (conceptual-to-empirical), starting the taxonomy definition process with an abstract or theoretical foundation and then deriving the structure through deduction. It identifies taxonomy dimensions and characteristics by a logical process derived from a sound conceptual or theoretical foundation, not based on empirical data (BAILEY, 1994). This approach is indicated when research significantly understands the target domain, but little data is available (NICKERSON; VARSHNEY; MUNTERMANN, 2013). The taxonomy definition method consists of four main phases: Planning, Identification, Design and Construction, and Validation (see Figure 16).

The planning phase consists of a definition of the taxonomy's context and initial setting. The first step is determining the meta-characteristics, the most comprehensive characteristic that will be the basis for choosing taxonomy elements (NICKERSON; VARSHNEY; MUNTERMANN, 2013). This process consists of defining the new taxonomy's knowledge area; objectives, scope, and subject; classification structure (tree, hierarchy, paradigm, or facet-based) and procedures (qualitative or quantitative); and data sources and data collection methods adopted. The second step is determining the objective and subjective ending conditions, which are rules used to determine when to stop the interactive design and construction process. We adopted the Nickerson, Varshney, and Muntermann (2013) 's ending conditions, as presented in Table 6.

In the identification phase, we collected relevant data to define the new taxonomy by conducting a systematic mapping study addressing the last ten years of leading CSCW publications (as presented in Chapter 3). Then, relevant terms were collected, and redundancies and inconsistencies were identified and removed using a terminology control process.

As presented in the literature (see Section 2.8), some strategies, frameworks, guidelines, design requirements, or groupware heuristics were applied during the development and evaluation of the awareness support, including awareness checklists (ANTUNES et al., 2014; ESPIRITO SANTO et al., 2018); awareness requirements (MANTAU; BERKENBROCK; BERKENBROCK, 2017); usability groupware heuristics(DE ARAÚJO et al., 2014); frameworks or taxonomies (GALLARDO; MOLINA, et al., 2011; SOUZA; BARBOSA, 2015; COLLAZOS et al., 2019; NIEMANTSVERDRIET et al., 2019). We will use these works as a starting point, grouping the literature into a hierarchical view of awareness support, represented by awareness dimensions, design categories, and support mechanisms.

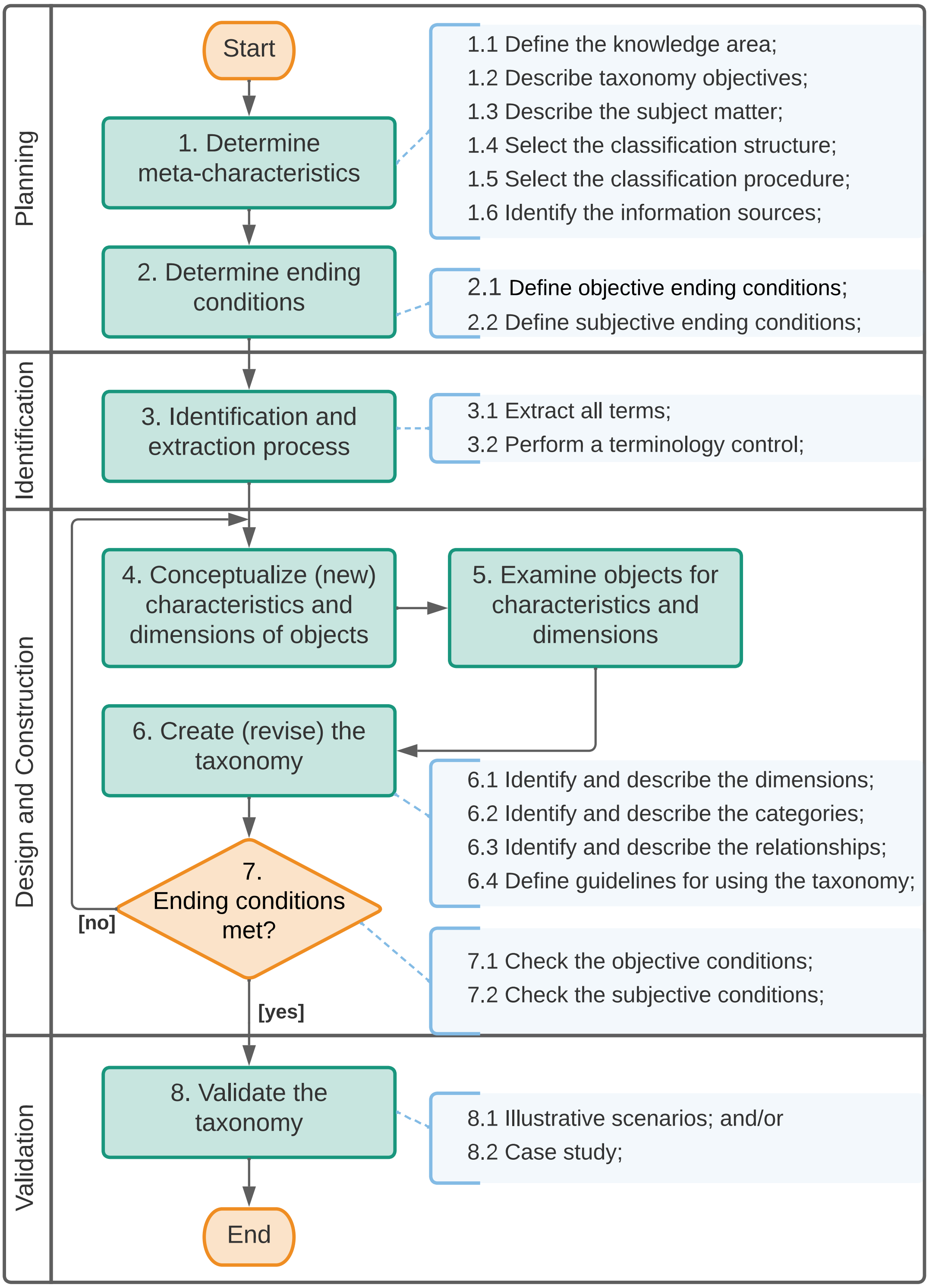


Figure 16 – Our taxonomy definition method

The design and construction phase consists of performing three main steps iteratively until all ending conditions are met. The first step is to examine if new characteristics may fit into existing or new dimensions may be conceptualized. In the next step, the researcher examines empirical cases using unique features and dimensions to determine their usefulness in classifying objects. Then, a new version of the taxonomy is defined, identifying and describing its dimensions, characteristics, relationships, and supplementary information (activities 6.1 to 6.4).

Table 6 – Objective and subjective ending conditions (NICKERSON; VARSHNEY; MUNTERMANN, 2013)

objective ending conditions	representative	→ All objects or a representative sample of objects has been examined
	mapping	→ At least one object is classified under every characteristic of every dimension
	unique	→ Each dimension, characteristic, or object is unique and not repeated (mutual exclusivity)
	complete	→ No dimensions, characteristics, or objects were added, merged, or split in the last iteration
subjective ending conditions	concise	→ The dimensions are meaningful without being unwieldy or overwhelming
	robust	→ The dimensions and characteristics provide sufficient differentiation among objects
	comprehensive	→ All dimensions of the objects of interest were identified
	extensible	→ A new dimension or characteristic can be easily added
	explanatory	→ The dimensions and characteristics explain about all identified object

We used a phenetic analysis during the design and construction steps, classifying elements by similarity (NICKERSON; VARSHNEY; MUNTERMANN, 2013). We identified different awareness characteristics/elements presented and clustered them into similar groups.

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. Finally, the validation phase ensures that the taxonomy will be helpful for users to achieve their goals and strengthens their reliability and usefulness (USMAN et al., 2017). Illustrative scenarios and case studies are the two most used methods in the literature for the validation of taxonomies (USMAN et al., 2017; SZOPINSKI; SCHOORMANN; KUNDISCH, 2019) and are indicated when a conceptual approach is adopted (SZOPINSKI; SCHOORMANN; KUNDISCH, 2019).

4.1 THE AWARENESS TAXONOMY

Our taxonomy resulted in three main awareness dimensions in collaborative applications: collaboration, workspace, and contextual awareness. These dimensions consist of a 4-level hierarchical representation structure, containing the awareness dimension, their design categories, design elements/ awareness mechanism involved, and the 5W+1H framework correspondence according to Gutwin and Greenberg (2002).

After the design and construction process, we found 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 our 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 Persona dimension is related to the owner/persona and indicates to whom the awareness information belongs (*who?*). This dimension allows classifying awareness elements among individuals, other participants, the group as a whole, or groupware/system perspective. The Boundary dimension indicates if awareness elements belong to the physical or virtual context of the collaborative environment (*where?*); thus, this dimension applies to details contained in the contextual awareness view.

The third and central additional dimension is historical awareness (*when?*). It represents the temporal information (past, present, and future) carried out during collaborative work, whether situational, contextual, or workspace information. In the literature, we can find historical awareness as a common type of awareness information; however, in our understanding, this approach is incompatible with its real nature. The historical perspective is broader and encompasses all other existing awareness elements in more categories.

4.2 COLLABORATION AWARENESS CATEGORY

Collaboration awareness refers to the notion of group availability, its structure, and interaction aspects. It was categorized into five design categories and contained 23 design elements, as presented in Figure 17. Table 53 in Appendix B presents the detailed structure of this vision, relating the design categories, its elements, and the relationship with the 5W+1H framework.

Table 7 contains the multidimensional relationship between this awareness view structure and the persona and historical awareness dimensions. This table correlates awareness elements based on these two additional views, considering the features described in the literature. For example, the Identity element was used in the past, present, and future (historical perspective) and the individual, other participants, or group as a whole (persona perspective). However, applications can represent this information from another temporal perspective, thus recording the changes in this aspect over time (past and present).

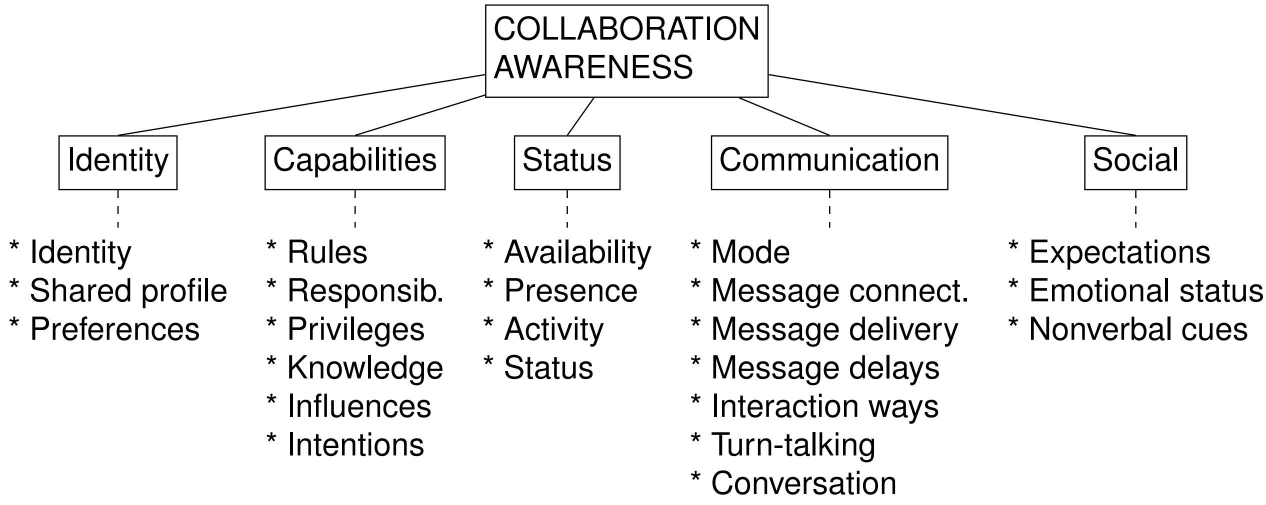


Figure 17 – Collaboration awareness category

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 rules and to design the cooperative work. It contains six design elements: *rules* 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.

Table 7 – Collaboration awareness dimensions

		Persona dimension				Historical awareness		
Design category	Design element	Individual	Other participants	Group as a whole	Groupware/System	When (past)	When (present)	When (future)
Identity	Identity	X	X	X	-	-	X	-
	Shared profile	X	X	-	-	-	X	-
	Preferences	X	X	X	-	-	X	-
Capabilities	Rules	X	X	X	X	X	X	X
	Responsibilities	X	X	X	-	X	X	X
	Privileges	X	X	X	-	X	X	X
	Knowledge	X	X	X	-	X	X	X
	Influence	X	X	X	-	X	X	X
	Intentions	X	X	X	-	X	X	X
Status	Availability	X	X	X	-	X	X	X
	Presence	X	X	X	-	X	X	X
	Activity level	X	X	X	-	X	X	X
	Status	X	X	X	X	-	X	-
Communication	Mode	X	X	-	-	-	X	-
	Network connectivity	-	-	-	X	-	X	-
	Message delivery	-	-	-	X	-	X	-
	Message delays	-	-	-	X	-	X	-
	Interaction ways	X	X	X	-	-	X	-
	Turn-taking	X	X	X	-	-	X	X
	Conversation	X	X	X	-	X	X	X
Social	Expectations	X	X	X	-	X	X	X
	Emotional status	X	X	X	-	X	X	X
	Other nonverbal cues	X	X	X	-	X	X	X

4.3 WORKSPACE AWARENESS CATEGORY

Workspace awareness 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 Figure 18. Table 54 in Appendix B presents the structure of this vision, its design categories and elements, and the relationship with the 5W+1H framework.

Table 8 contains the multidimensional relationship among this awareness view structure and Persona and Historical awareness dimensions (similar to one presented in the collaborative awareness view). These awareness elements can also be considered in a physical or virtual context.

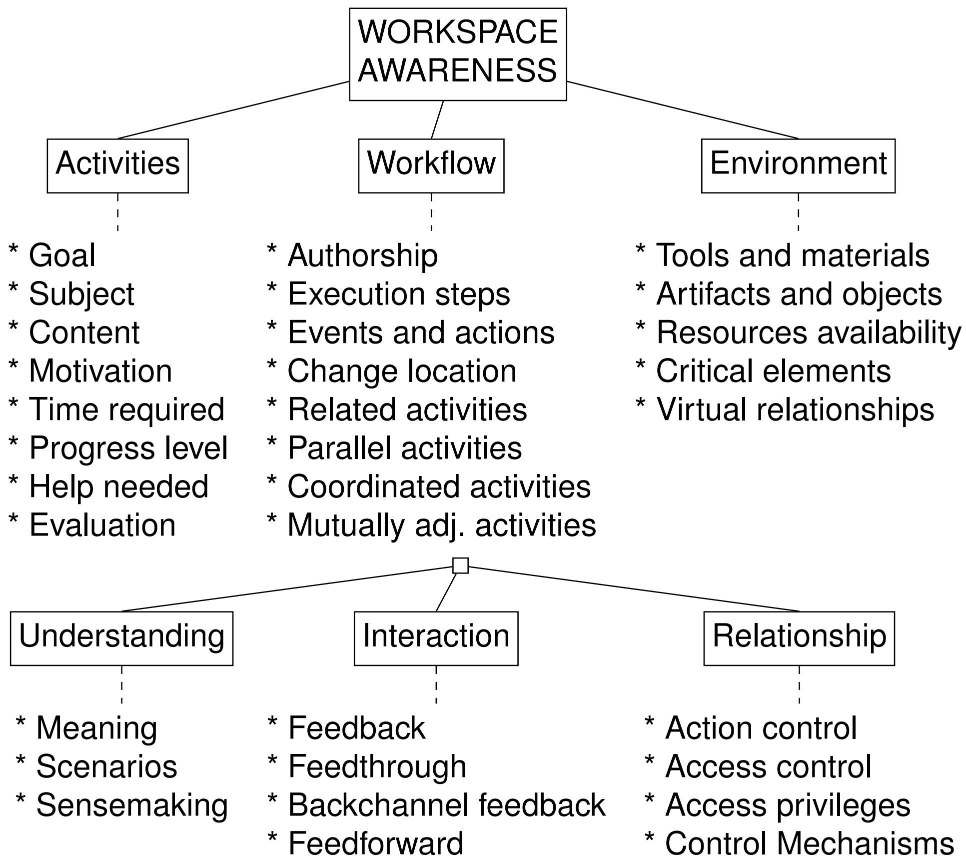


Figure 18 – Workspace awareness category

Table 8 – Workspace awareness dimensions

		Persona dimension				Historical awareness		
Design category	Design element	Individual	Other participants	Group as a whole	Groupware/System	When (past)	When (present)	When (future)
Activities	Goal	X	X	X	-	X	X	X
	Subject	X	X	X	-	X	X	X
	Content	X	X	X	-	X	X	X
	Motivation	X	X	X	-	X	X	X
	Time required	X	X	X	-	X	X	X
	Progress level	X	X	X	-	X	X	X
	Help needed	X	X	X	-	X	X	X
	Evaluation	X	X	X	-	X	X	X
Workflow	Authorship	X	X	X	-	X	X	X
	Execution steeps	X	X	X	-	X	X	X
	Events and actions	X	X	X	-	X	X	X
	Change location	X	X	X	-	X	X	X
	Related activities	X	X	X	-	X	X	X
	Parallel activities	X	X	X	-	X	X	X
	Coordinated activities	X	X	X	-	X	X	X
	Mutually adjusted activities	X	X	X	-	X	X	X
Environment	Tools and materials	X	X	X	-	X	X	X
	Artifacts and objects	X	X	X	-	X	X	X
	Resources availability	X	X	X	X	X	X	X
	Critical elements	X	X	X	-	X	X	X
	Virtual relationships	X	X	X	-	X	X	X
Understanding	Meaning	X	X	X	-	X	X	X
	Scenarios	X	X	X	-	X	X	X
	Sense-making	X	X	X	X	X	X	X
Interaction	Feedback	X	-	-	-	-	X	-
	Feedthrough	-	X	X	-	-	X	-
	Backchannel feedback	X	-	-	-	-	X	-
	Feedforward	X	-	-	-	-	X	-
Relationship	Action control	X	X	X	-	-	X	-
	Access control	X	X	X	-	-	X	-
	Access privileges	X	X	X	-	-	X	-
	Control mechanisms	X	X	X	-	-	X	-

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 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; *Backchannell 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).

4.4 CONTEXTUAL AWARENESS CATEGORY

Contextual Awareness 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 Figure 19. Table 55 in Appendix B represents the structure of this vision, relating its design categories and elements and the relationship with the 5W+1H framework.

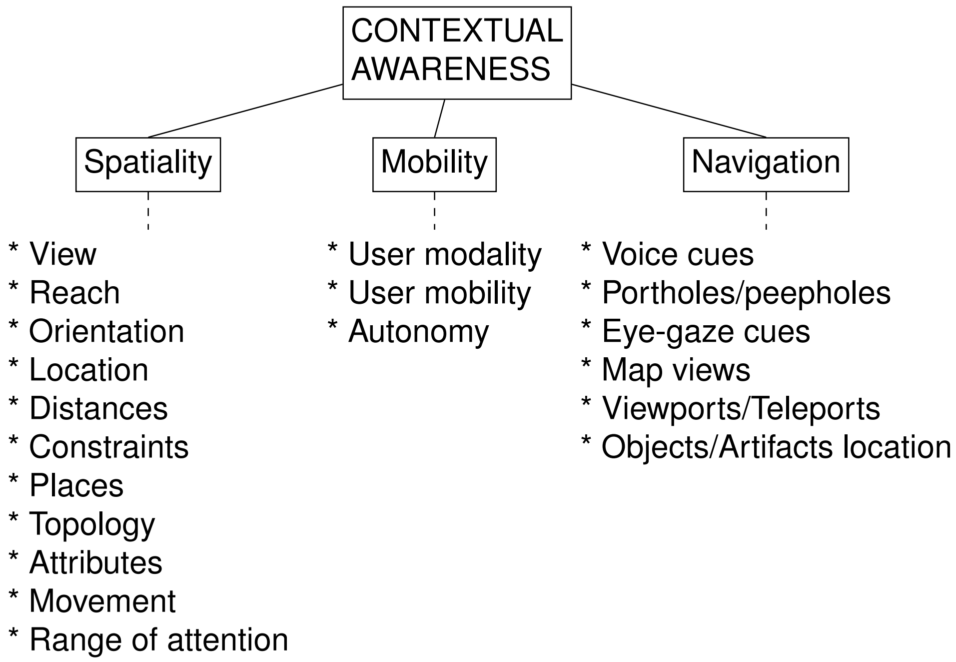


Figure 19 – Contextual awareness category

Table 9 contains the multidimensional relationship among this awareness view structure and Boundary, Persona, and Historical awareness dimensions.

Table 9 – Contextual awareness dimensions

		Boundary dimension		Persona dimension				Historical awareness		
Design category	Design element	Physical	Virtual	Individual	Other participants	Group as a whole	Groupware/System	When (past)	When (present)	When (future)
Spatiality	Location	X	X	X	X	X	-	X	X	X
	Distances	X	X	X	X	-	-	X	X	X
	Constraints	X	X	X	X	X	X	-	X	-
	Places	X	X	X	X	X	-	X	X	-
	Topology	X	X	-	-	-	X	-	X	-
	Attributes	X	X	X	X	X	X	X	X	X
	View	X	X	X	X	X	-	-	X	-
	Reach	X	X	X	X	X	-	-	X	-
	Orientation	X	X	X	X	X	-	X	X	X
	Movement	X	X	X	X	X	-	X	X	X
	Range of attention	X	X	X	X	X	-	X	X	X
Mobility	User modality	X	-	X	-	-	-	-	X	-
	User mobility	X	-	X	-	-	-	-	X	-
	Autonomy	X	X	-	-	-	X	-	X	-
Navigation	Voice cues	-	X	X	X	-	-	-	X	-
	Portholes/peepholes	-	X	X	X	X	-	-	X	-
	Eye-gaze cues	-	X	X	X	X	-	-	X	-
	Map views	-	X	X	X	X	-	-	X	-
	Viewports/Teleports	-	X	X	X	X	-	-	X	-
	Artifacts location	X	X	X	X	X	-	-	X	-

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 eleven design elements: *Location* of each participant over time, whether the user is in same physical place as another; *Distances* of user in relation to others; *Constraints* imposed by physical environment where it is used (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 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 in regard to other users; and *Range of attention* when performing activities.

Mobility 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.

4.5 TAXONOMY EVALUATION

The resulting taxonomy must be evaluated for its usefulness (NICKERSON; VARSHNEY; MUNTERMANN, 2013). To assess the usefulness of taxonomy, we consider three fundamental aspects: purposeful, unambiguous determination, and applicability (STRASSER, 2017), as presented in Figure 20.

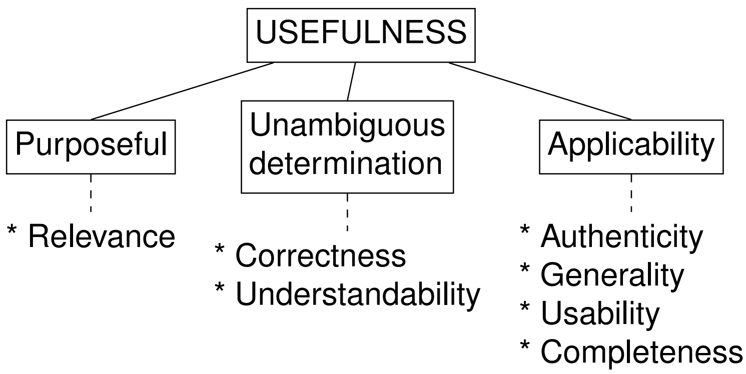


Figure 20 – Usefulness evaluation criteria

Purposeful is the significance and objectivity of the model and its elements (STRASSER, 2017), and it is related to the taxonomy relevance. Relevance is archived when all statements in the representation are relevant to the problem (RITTGEN, 2010).

Unambiguous determination is the ability to represent the elements and characteristics in a clear, concise, and unambiguous way (STRASSER, 2017), and it is related to the taxonomy correctness and understandability. Taxonomy met correctness when all statements in the representation are correct (RITTGEN, 2010); understandability when the reference model's purpose, concepts, and structure are clear to the users (MATOOK; INDULSKA, 2009).

Applicability seeks to assess its practical purpose and usefulness for classifying, differentiating, and comparing objects (STRASSER, 2017), and it is related to the taxonomy authenticity, generality, usability, and completeness. Authenticity evaluates if the representation gives a true account of the domain (RITTGEN, 2010); generality indicates if the reference model is usable in different cases (MATOOK; INDULSKA, 2009); usability consists of identifying if users can easily operate, implement, and apply the reference model (MATOOK; INDULSKA, 2009); and completeness check if all the components of the reference model are present under a predefined scope (MATOOK; INDULSKA, 2009) and the representation contains all statements about the domain that are correct and relevant (RITTGEN, 2010).

To assess the aspects of purposeful and unambiguous determination, we used four classifications described in the literature and correlated them with our developed taxonomy. First, we correlate our taxonomy and its elements with the literature (ANTUNES et al., 2014; ESPIRITO SANTO et al., 2018; GALLARDO; BRAVO; MOLINA, 2018; NIEMANTSVERDRIET et al., 2019). Next, we look at the significance, objectivity, clear, concise, and unambiguous properties.

All design categories established in our taxonomy have unique and representative design elements that can be related to those presented by existing classifications. The Tables 56, 57, and 58 in Appendix B correlates the three awareness views of our taxonomy, respectively, the collaboration, workspace, and contextual view, with existing classifications presented in the state of the art of the collaborative systems literature and their elements.

4.6 CONSIDERATIONS

Awareness is a multifactorial 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 novices in awareness or collaborative systems design (NIEMANTSVERDRIET et al., 2019); they must reinvent awareness from their own experience of what it is, how it works, and how it is used (COLLAZOS et al., 2019).

Our approach presented a strategy broadly considering awareness, moving us towards a global model to help designers understand, develop, and support awareness in collaborative applications. In this research, we use the awareness taxonomy to develop a novel assessment strategy to enable collaborative applications to achieve the collaboration aspects necessary for cooperative work by assessing the awareness support. We also believe new strategies or approaches can emerge based on this taxonomy, whether in design, development, or evaluation.

Our taxonomy could represent the awareness elements identified in the literature through our systematic mapping into three main awareness dimensions: collaboration, workspace, and contextual perspectives. This aligns with the need to organize and simplify the representations of awareness defined in the literature. We could map the awareness elements described in other literature models to our taxonomy representation through the illustrative scenario approach. This property allows us to classify elements of awareness foreseen in such representations.

Furthermore, the persona, boundary, and historical additional dimensions of our taxonomy provide flexibility and allow awareness elements to be classified according to the owner of the information, the physical/virtual boundary to which it belongs, and the temporal label being considered. This innovative multidimensional framework for awareness mechanisms simplifies existing representations that work with these perspectives. We observed in the literature a difficulty in conceptualizing/representing the awareness information over the persona, boundary, and historical dimensions, where the solution adopted is to establish a new awareness element whenever the concept is involved in one of these perspectives.

4.6.1 Related publications

The results presented in this chapter were published in:

- MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. Towards an Awareness Taxonomy. In: IEEE. 2022 IEEE 25th International Conference on Computer-Supported Cooperative Work in Design (CSCWD). IEEE, 2022. P. 495–500. DOI: 10.1109/cscwd54268.2022.9776129.

5 AWARENESS ASSESSMENT MODEL

The Awareness Assessment Model is explicitly developed for evaluating collaborative systems. It measures their quality by analyzing the awareness information provided by the application. 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.

5.1 MODEL ASSUMPTIONS

In our assessment model, we assume the following assumptions:

i) 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;

ii) Collaboration results from the participant's understanding/consciousness. The consciousness allows individuals to project their actions;

iii) 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.

5.2 MODEL OVERVIEW

The Awareness Assessment Model comprises the Awareness Assessment Process and the Conceptual View (see Figure 21).

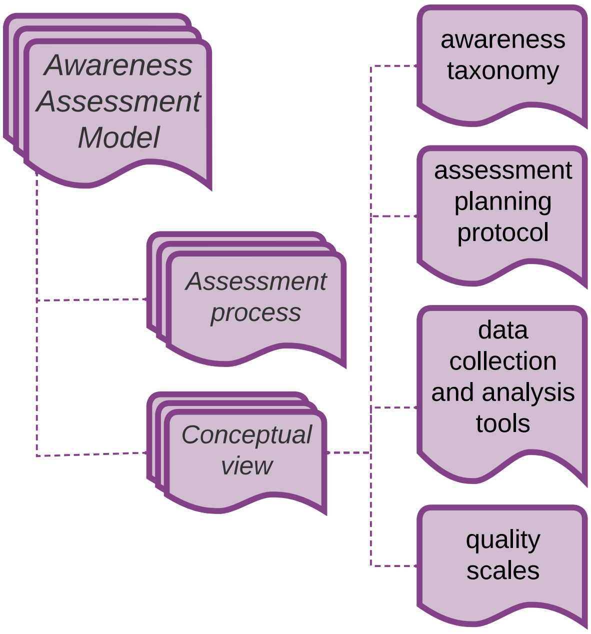


Figure 21 – Awareness Assessment Model overview

5.3 THE AWARENESS ASSESSMENT PROCESS

Our awareness assessment process is based on a set of HCI guidelines (BARBOSA; SILVA, 2010; ROGERS; SHARP; PREECE, 2013) and is inspired by the evaluation process defined by the standard ISO/IEC 25040:2011 (STANDARDIZATION, 2011).

The assessment process consists of three main phases: planning, execution, and reflection (see Figure 22). This process is performed by the researcher/examiner, who evaluates the collaborative interfaces by analyzing the awareness information provided by the application. This process involves the participation of a sample of target users, in which the environment is evaluated using data collection and analysis tools.

Phase 1 - Planing. It refers to activities related to assessment planning and involves three basic steps: determine the assessment objectives, the assessment scope, 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 (BASILI, 1992): 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 or environment in which the evaluation is performed.

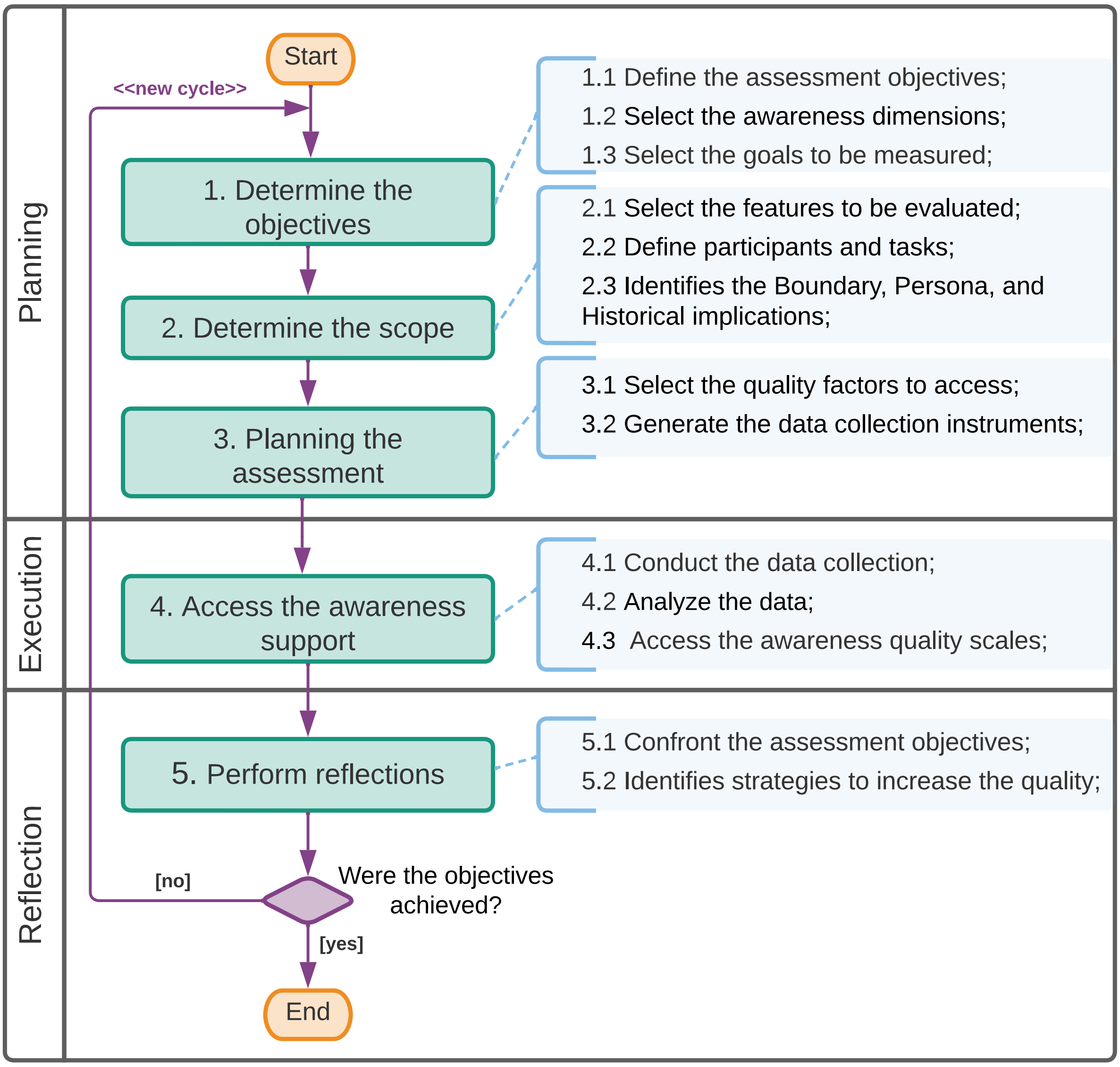


Figure 22 – Our awareness assessment process

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 dimensions, allowing 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, allowing the model's flexibility to address the application's relevant aspects.

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 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, it is vital to clarify who is involved and what their tasks are in the environment.

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, 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 provided by the target environment is reached through 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 not met, the examiner determines strategies to increase the quality indicators of the awareness mechanisms, 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.

5.4 THE CONCEPTUAL VIEW

The Conceptual View consists of a framework, and it 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 (Chapter 4);

ii) 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 (Section 5.5);

iii) The data collection and analysis tools present a set of support artifacts for conducting the collection and compilation of data obtained by interventions (Section 5.6);

iv) The awareness assessment quality scales and measurement are useful for analyzing the results obtained through assessment instruments and classifying the collaborative environment at a quality level through the participants' perception (Section 5.7).

5.5 ASSESSMENT PLANNING PROTOCOL

The planning protocol is a form that helps in planning the evaluation. It consists of three main pieces of information: determining the intervention's objectives, scope, and life-cycle. The planning protocol template example presented in Table 10 details the procedures performed in the planning stage of our assessment process.

Table 10 – Planning protocol template

1. Determine the objectives	Step 1.1. Define the assessment objectives regarding the object of study, purpose, perspective, and context. → *Object of study:* → *Purpose:* → *Perspective:* → *Context/environment:* Step 1.2. Select the awareness dimensions. Identify the related awareness dimensions that will be considered in the assessment. → *Awareness dimensions:* Step 1.3 Select the goals to be measured. We select which design categories are relevant in the collaborative environment for each awareness dimension. → *Goals:*
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. → *Features:* Step 2.2 Define the participants involved and the tasks that must be carried out within the collaborative environment. → *Participants:* → *Tasks description:* 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. → *Implications:*
3. Planning the intervention	Step 3.1. Select the additional factors to access (like demographic, usability, and UX). → *Additional factors:* Step 3.2. Generate the data collection instrument considering the activities 1.2, 1.3, and 2.3.

5.6 DATA COLLECTION AND ANALYSIS TOOLS

The awareness assessment model contains a set of data collection instruments that are applied through assessment questionnaires. In the full version of the model, we considered adopting 75 specific awareness assessment items identified in our taxonomy divided into ten ten awareness questionnaires. It was developed based on a multidimensional perspective represented by three main awareness dimensions as presented in Chapter 4.

The ten questionnaires were balanced for each awareness dimension. The assessment items were developed using the guidelines presented by BASILI (1992), Wohlin et al. (2012), Wohlin (2014), and DeVellis (2016). 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).

Tables 11 to 14 present the complete version of the questionnaire. The questions were composed by combining the following structure: the component of the sentence (subject + predicate) present in Table 11 combined with the correspondent complement of each evaluation item (Table 12, 13, or 14, according to the awareness dimensions evaluated).

All assessment items were composed similarly. For example, the assessment item Q1 is the combination of component s1 (Table 11) + questionnaire item complement "Goal" (Table 12); thus, we composed the questionnaire item Q1 into *"the collaborative environment allows me to identify the purpose of the tasks performed"*.

Table 11 – 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"

The awareness views in the questionnaire constitute the complete version of the design categories and awareness elements identified in our taxonomy. We designed our approach to provide flexibility in the application of the instrument by the evaluators through the prior selection of categories/design elements to apply.

Table 12 – Assessment items: workspace awareness dimension

Awareness Taxonomy		Item	Questionnaire items (object/complement)
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 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”

continues on the next page

Table 12: Assessment items: workspace awareness dimension (continuation)

	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 to access them”

Table 13 – Assessment items: collaboration awareness dimension

Awareness Taxonomy		Item	Questionnaire items (object/complement)
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 status/situation of each participant”
Provide communication	Mode	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 others”
	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 others”
Consider social	Expectations	Q53	s5 + “what are the expectations involving each participant”
	Emotional status	Q54	s5 + “the emotional status of each participant”
	Non-verbal cues	Q55	s5 + “the availability of non-verbal information for communication”

Table 14 – Assessment items: contextual awareness dimension

Awareness Taxonomy		Item	Questionnaire items (object/complement)
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 way (e.g., a 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”

5.6.1 Balancing the questionnaire items

As proposed in our assessment process, we developed a flexible model that can be applied both in the full version and in a simplified view mode (e.g., selecting which dimensions and awareness assessment elements are applicable to assess the target environment). Thus, in the full mode, we considered adopting all 75 assessment items in the awareness taxonomy. To reduce the number of assessment items for each participant, we used the balanced incomplete block design approach (HINKELMANN; KEMPTHORNE, 2005). This approach is generally used when the number of treatments (assessment items, in our case) is hard to apply.

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; 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 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 15 presents the configuration of the treatments (t) and blocks of questionnaire items (b). A complete example of the ten assessment questionnaires is available in Appendix C, Figures 70 to 79.

Table 15 – 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)
$b1 = \{t1, t2\}$		$b6 = \{t3, t4\}$
$b2 = \{t3, t5\}$		$b7 = \{t2, t4\}$
$b3 = \{t1, t4\}$	...	$b8 = \{t4, t5\}$
$b4 = \{t1, t5\}$		$b9 = \{t1, t3\}$
$b5 = \{t2, t3\}$		$b10 = \{t2, t5\}$

5.7 AWARENESS MEASUREMENT AND ASSESSMENT QUALITY SCALES

The awareness measurement mechanisms and awareness quality scales aim to classify the collaborative environment at a quality level through the participants' perspective. To help the examiner and guide him/her through the awareness assessment process, we developed the data collection and analysis tools (Section 5.6).

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 Samejima (1969) and Baker and Kim (2004).

5.7.1 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 well responses in that category estimate the examinee's ability (BAKER; KIM, 2004). Our model assumes 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 15).

$$I_i(\theta) = \sum_{k=0}^{n} \frac{[P^{*\prime}_{i,k-1}(\theta) - P^{*\prime}_{ik}(\theta)]^2}{P^{*}_{i,k-1}(\theta) - P^{*}_{ik}(\theta)} \tag{15}$$

with,

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

- 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 Baker and Kim (2017) (Equation 16). Where m is the amount of applicable items of goal g; $I_j(\theta)$ is the item's information for each applicable goal item j.

$$GI(\theta) = \sum_{j=1}^{m} I_j(\theta) \tag{16}$$

- The awareness dimension's information $AI(\theta)$, for each applicable awareness dimension d, that is calculated considering all applicable goal g using the test information function presented by Baker and Kim (2017) (Equation 17). Where o is the number of related goals g; $GI_j(\theta)$ is the amount of information for each applicable goal g.

$$AI(\theta) = \sum_{l=1}^{o} GI_l(\theta) \tag{17}$$

It is essential to highlight that the Equations 15, 16, and 17 calculate the information scores from a single participant viewpoint (BAKER; KIM, 2017), 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.

5.7.2 Assessment Quality Scales

The assessment quality scales have been developed adopting the Item Response Theory (IRT) statistical technique, as presented by Baker and Kim (2017). The IRT refers to a family of mathematical models that relate observable variables (e.g., questionnaire items) and hypothetical unobservable traits or aptitudes (e.g., awareness quality). Thus, a stimulus (item) is presented to the subject, and he 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 (PASQUALI, 2020).

The IRT model is built by executing scripts in R source using the MIRT package (a multidimensional Item Response Theory package for the R environment) (CHALMERS, 2012). All sources and related materials are in the repository (MANTAU; BENITTI, 2023).

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; 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 4 (see Section 2.6.2.2).

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 describes how well an item can discriminate (differentiate) the participants concerning 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 interpreted following Baker and Kim (2017). A measurement instrument item is satisfactory in a measurement scale if the discrimination value $a \geq 0.65$, as presented in Table 16. 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 awareness quality levels, we considered the probability parameter $P_{i,k}(\theta) \leq 0.5$ and scale $(0, 1)$ (ANDRADE; TAVARES; VALLE, 2000). 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, parameters a and b vary between $[-4.0, +4.0]$. The most appropriate values of a would be greater than 0.65.

Table 16 – Item discrimination parameter values (BAKER; KIM, 2017)

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

The IRT calculates and positions a participant's score on the defined ability scale. However, we are interested in classifying the collaborative environment; thus, we calculate the average of the provided scores of all participants involved.

Based on the positioning of items throughout the scale, three levels of quality are defined: low quality ($\theta < -1$), good quality ($-1 \leq \theta \leq 1$), and excellent quality ($\theta > 1$). We assume a coverage interval of $[-4.0, +4.0]$ to construct our awareness support scale. The scale $[-4.0, +4.0]$ is commonly used in the IRT models (DEVELLIS, 2016).

In our 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. In other words, we represent the expected ability intervals in which participants present a certain probability P_i of selecting each response category presented, that is, the probable intervals $P_i(\theta)$ that participants are most likely to correctly identify/understand the awareness mechanism in the evaluated interface.

This gradual organization allows prioritizing 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 workspace, collaboration, and contextual awareness mechanisms.

5.8 CONSIDERATIONS

Collaborative environment assessment is dynamic and involves considering several factors throughout the assessment process. We understand that for an appropriate awareness assessment, it is essential to consider aspects of the participant himself because, from an individualized point of view, each participant is a unique actor, whether in the identification, recognition, and projection of his actions through the resources or mechanisms available for interaction with others.

To assess awareness, the indissociable relationship between what awareness is and how it supports each collaboration facet is maintained. Just as we cannot adequately understand what awareness represents in a collaborative environment without considering the surrounding context, different people have different knowledge, skills, expectations, and interaction needs. Thus, the evaluation process must comply with this premise, incorporating, as much as possible and, as an active part of the model, a broader awareness perspective corresponding to each user's point of view.

The developed assessment model seeks to bring the evaluation of awareness support closer to each user's perspective. We establish a theoretical framework through the conceptual vision comprising three awareness dimensions, related design categories, and awareness elements/mechanisms. The conceptual view also describes a set of artifacts that guide the entire evaluation process, from data collection instruments, analysis tools, and awareness scale construction and interpretation. The support scale highlights the quality of the collaborative environment by considering the participant's ability to identify each awareness element.

Although awareness is deeply linked to each participant's context and knowledge, our model establishes various design categories and associated awareness elements that can be incorporated into customized analysis instruments. In addition, the proposed evaluation model can be used both in a full version, considering all awareness mechanisms, design categories, and dimensions, or in a simplified or partial evaluation view, where a subset of these items, more relevant for the intended assessment, can be adopted. By selecting the more appropriate awareness mechanisms in each evaluation, we believe the scale results and participants' ability to identify this information are more reliable to the target context.

An awareness assessment process was defined, inspired by HCI assessment guidelines. It incorporates an iterative and incremental approach, where assessment cycles are carried out, and the process feeds back over findings and reflections provided at each new cycle.

5.8.1 Related publications

The assessment model presented in this chapter was published in:

- MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. The Awareness Assessment Model: measuring the awareness and collaboration support over the participant's perspective. In: Anais do XVIII Simpósio Brasileiro de Sistemas Colaborativos, SBSC. SBC, 2023. P. 30-43.
- MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. The Awareness Assessment Model: Measuring Awareness and Collaboration Support Over Participant's Perspective. In: Universal Access in the Information Society (UAIS). DOI: 10.1007/s10209-024-01110-5.
- MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. The Awareness Assessment Model repository. Version 1.0. Zenodo, Aug. 2023. DOI: 10.5281/zenodo.8298950.

Part III

Model Validation

6 THE ASSESSMENT MODEL VALIDATION

The validity of an item or instrument refers to the fact that it is related to what you want to measure (PASQUALI; PRIMI, 2003). There are three types of validity: criterion, content, and construct validity (PASQUALI; PRIMI, 2003; RICHARDSON, 2017).

According to Richardson (2017), 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 (NUNNALLY; BERNSTEIN, 1993).

The 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 (RICHARDSON, 2017).

The construct validity concerns the validation of a theory, which is reflected in a given instrument (RICHARDSON, 2017). Nunnally and Bernstein (1993) defines construct validity as the extent to which the set of items measures a theoretical latent trait. Construct validity is 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. According to Pasquali and Primi (2003), the construct validity can be analyzed from several angles, from Classical Measurement Theory (TCM) and Item Response Theory (IRT).

Construct validity can also be defined as the extent to which a set of items, or tests, measures a latent trait (PASQUALI; PRIMI, 2003). In our model, the latent trait is the support for awareness and collaboration provided by the collaborative environment. Urbina (2009) list as procedures that identify 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 (BEECHAM et al., 2005). 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 (detailed in Section 2.7). 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. The results of this step are presented in Chapter 7.

After this refinement, we started planning and executing a set of case studies (WOHLIN et al., 2012; YIN, 2009) through a large-scale evaluation of our assessment model. This approach evaluates the proposed model's reliability, validity, and dimensionality. We pooled the data as a single sample for data analysis for each case study. We evaluated the proposed model on reliability and dimensionality (detailed in Sections 2.7.1 and 2.7.2).

Data on reliability and construct validity were analyzed following the definition of (TROCHIM; DONNELLY, 2001) and the scale development guide (DEVELLIS, 2016). We considered internal consistency through the Cronbach alpha coefficient (CRONBACH, 1951) combined with IRT parameters for the reliability measurement. IRT allows us to evaluate the quality of the assessment items through θ, discrimination, and difficulty parameters. The test information function $I(\theta)$ was also used to calculate standard error and reliability (DEMARS, 2010).

Exploratory Factor Analysis (EFA) and Confirmatory factor analysis (CFA) were applied to test dimensionality (HAIR et al., 2009; PASQUALI; PRIMI, 2003; IZQUIERDO; OLEA; ABAD, 2014).

6.1 CASE STUDY SCENARIOS

We developed three case studies.

In the first scenario (described in Chapter 8), we evaluated a specific collaborative environment: The Moodle platform. This scenario was designed to refine the model and adjust/adequate the artifacts. For this reaction, we selected the evaluation of a single collaborative environment, considering a small sample of participants.

In the second scenario (described in Chapter 9), we evaluated a set of collaborative videoconferencing environments. Here, we evaluated a set of collaborative applications to verify the behavior of the evaluation model in each.

In the third scenario (described in Chapter 10), we evaluated a set of collaborative text editing environments. At this point, we expose the model artifacts to the appreciation of HCI and collaborative system examiners to verify the adequacy of the process, its activities, and related artifacts to evaluate awareness support in collaborative environments. First, examiners adopted the model's artifacts and process to evaluate the awareness support in general-purpose collaborative office tools (e.g., most common text editing tools, spreadsheets, and document managers). Second, researcher observations and questionnaires assessed the usefulness of the model's conceptual view artifacts and evaluation process.

In both 1st and 2nd case study scenarios, we applied the complete evaluation model, considering all awareness dimensions and design categories; therefore, the environments were analyzed using the 75 assessment items described in the model. In the 3rd scenario, each group of examiners selected a subset of assessment items presented in the model and prepared their assessment targeted to the target environment, making it possible to evaluate the model's behavior when partially used.

6.2 THE GLOBAL AWARENESS SCALE

In Part IV Chapter 11, we present the global awareness scale based on data obtained through scenarios of case studies 1 to 3 (Chapters 8 to 10). We assume the estimation of both participants' abilities and items' parameters as a calibration strategy and the IRT multi-group estimation method (CHALMERS, 2012). To calibrate the model and generate the global awareness scale, we considered each of the scenarios as a distinct evaluation group: case study 1 (group 1), case study 2 (group 2), and case study 3 (groups 3 to 9).

In the global awareness scale scenario, two or more groups take two or more tests, which are only partially different (with some common items). In this configuration, the common elements between different tests allow all parameters to be on the same scale at the end of the estimation processes. As a set of elements connects the different populations, it is possible to make comparisons and construct a global scale (ANDRADE; TAVARES; VALLE, 2000).

7 EXPERT PANEL VALIDATION

When a new evaluation model emerges, the first concern is to check whether the model allows measuring the target; in our case, the latent trait of the awareness support. In other words, we need to validate whether the assessment instrument (especially its items) is applicable or representative of different situations (RICHARDSON, 2017). This analysis can be done by adopting the expert panel technique (BEECHAM et al., 2005).

Experts discuss the pretended object and make recommendations; this approach aims to hear expert viewpoints to decide on recommendations or courses of action concerning an issue or proposal. According to Hakim (1987), small samples can be used to develop and test explanations and to gain expert feedback to evaluate and support model development. Some previous works describe the use of small samples of experts to gain feedback to evaluate and support model development: software quality evaluation (ROSQVIST, 2003); prevent requirements defects (LAUESEN; VINTER, 2001); software process improvements (DYBA, 2000); and software requirements analysis (EL EMAM; BIRK, 2000).

To improve the assessment model, we expose the model artifacts to the appreciation of HCI and collaborative systems experts using the expert panel approach (BEECHAM et al., 2005). 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. After this refinement, we reviewed the exposed artifacts; then, we started the large-scale model evaluation process by planning and executing a set of case studies (described in Chapters 8 to 10) to access the model's construct validity.

From the researchers' perspective, the expert panel approach aims to analyze the usefulness aspects, namely, clarity, relevance, consistency, and completeness of the measurement instrument items. The usefulness is related to the purposeful, unambiguous determination and applicability aspects (NICKERSON; VARSHNEY; MUNTERMANN, 2013) (similarly as presented in Section 4.5, Fig 20).

Purposeful refers to the significance and objectivity of the model and its elements. Unambiguous determination refers to 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.

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 (STRASSER, 2017). This approach also allows reflecting on the current state of research on an object (KHALILIJAFARABAD; HELFERT; GE, 2016), discovering similarities and differences between studies on this type of object (AGOGO; HESS, 2018), and identifying potential research gaps (HUMMEL; SCHACHT; MAEDCHE, 2016).

Based on the Goal Question Metric approach (BASILI, 1992; VAN SOLINGEN; 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 17. In this step, we expose our assessment model to expert evaluation, such as awareness, collaborative systems, and HCI researchers, to identify its suitability to evaluate collaborative environments.

7.1 EXPERT PANEL RESULTS

We targeted experts from different backgrounds and audience groups, as recommended by Lauesen and Vinter (2001) and Kitchenham, Pfleeger, et al. (2002). We consider an expert in this study to be a researcher who has published widely in recognized journals in CSCW and HCI fields and has practical experience in evaluating collaborative environments and awareness.

We mailed 28 experts an invitation to validate assessment model artifacts, and five experts accepted (representing a take-up rate of 18%). The experts were mostly selected based on related publications identified in our systematic mapping and the researcher's contact list. As we cannot confirm the reason for the non-participation of the 23 invited experts, some likelihood of bias is present.

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). 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 items M1 to M7 received values over 3.5 (average 3.8). Figure 23 summarizes the obtained results.

The results obtained were positive in all facets of usefulness assessment. From the purposeful perspective, experts recognized the relevance and confirmed an adequate problem representation. From the unambiguous determination perspective, the results evidence the assessment statements' correctness. Finally, from the applicability perspective, the feedback pointed out that the representation provides an adequate account of the domain and that the assessment model artifacts can be usable in different cases, corroborating the authenticity and generality properties.

Table 17 – Expert panel questionnaire items

Validation aspect		Questionnaire item	Response
Demographic		Indicate your expertise in relation to the following:	
	D1	Awareness;	gradual scale (from novice to expert)
	D2	Collaborative systems;	
	D3	Human-Computer Interaction;	
Purposeful	M1	Relevance: all statements in the representation are relevant to the problem	
Correctness	M2	Correctness: all statements in the representation are correct	gradual scale (from strongly disagree to strongly agree)
	M3	Understandability: the purpose, concepts, and structure of the reference model are clear to the users	
Applicability	M4	Authenticity: the representation truly considers 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	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	

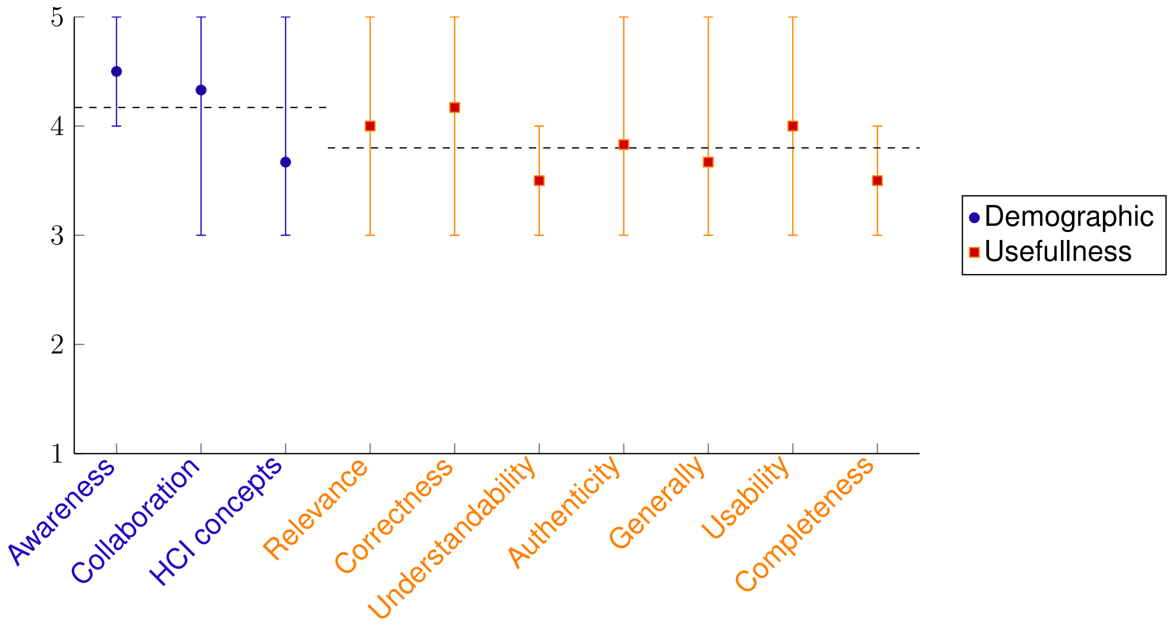


Figure 23 – Expert panel questionnaire results

In the applicability perspective, a central concern was related to the model's usability. Due to the natural complexity of evaluating awareness support, approaches from this perspective can be somewhat complex, and the model must abstract this complexity from the examiner as much as possible. On the other hand, from the point of view of the assessment items' usability, the majority agreed that target examiners could easily operate, implement, and apply the reference model; still, prior knowledge of the process and awareness support are recommended.

From the understandability and completeness aspects, the received feedback 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 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. At this stage, we also improved the assessment items' syntax by standardizing the assessment statements.

7.2 DISCUSSION

Similar to the procedure adopted in taxonomy assessment, the model artifacts need to be evaluated for their usefulness (NICKERSON; VARSHNEY; MUNTERMANN, 2013), and an expert panel scenario was used for evaluating it. In this approach, the assessment artifacts were applied in an expert panel scenario specially developed for the model evaluation to demonstrate the adequacy or usefulness of the artifact.

For the usefulness assessment, we considered purposeful, unambiguous determination and applicability characteristics (STRASSER, 2017). Experts' opinions pointed out the purposeful, unambiguous determination and applicability of model artifacts and their structure. Furthermore, we were able to access content validity: the artifacts of the assessment model (the questions or items) are representative, clear, and relevant, allowing them to be adopted in different situations. Considering that the model's artifacts represent evaluating collaborative environments, we can conclude that the model has content validity.

As the next steps towards validating the awareness assessment model, we developed a set of case studies involving the real application of the assessment instrument to enable the measurement of the construct's validity (Chapters 8 to 11).

7.2.1 Related publications

The results presented in this chapter were published in:

- MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. The Awareness Assessment Model: measuring the awareness and collaboration support over the participant's perspective. In: Anais do XVIII Simpósio Brasileiro de Sistemas Colaborativos, SBSC. SBC, 2023. P. 30-43.
- MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. The Awareness Assessment Model repository. Version 1.0. Zenodo, Aug. 2023. DOI: 10.5281/zenodo.8298950.

8 CASE STUDY 1: E-LEARNING SUPPORTING ENVIRONMENT

In the first case study scenario, we evaluate the collaboration aspects provided by the Moodle platform. Moodle is a collaborative learning platform designed to create learning environments for educators and students. For collaborative learning to occur satisfactorily in this environment, the application must provide awareness cues, like participant profiles, communication and interaction resources, and spaces to share artifacts, objects, and materials.

This case study aims to identify the level of collaboration support provided by the collaborative environment through a general evaluation of the awareness mechanisms presented in the participant's interface. In this scenario, we evaluate the Moodle platform's awareness support by assessing the commonly used Moodle resources, like chat, forum, assignments, lessons, and wiki (see the Moodle's assessment questionnaire available in Appendix C, Figures 68), and 70 to 79. The case study materials and the IRT dataset are available at Mantau and Benitti (2023).

8.1 PHASE 1: PLANNING

We planned the case study using the awareness assessment process (described in section 5.3). As a result of this step, the planning protocol artifact was created (Table 18), where we determined the intervention's objectives, scope, and life cycle.

Table 18 – Planning protocol (Moodle scenario)

Phase 1. Determine the objectives	Step 1.1. Define the assessment objectives. This activity defines the evaluation goal in terms of the object of study, purpose, perspective, and context. – Object of study: *Moodle environment;* – Purpose: *identify the level of collaboration support provided by the collaborative environment, through the evaluation of the awareness mechanisms presented;* – Perspective: *educators and (specially) students resources;* – Context or environment: *e-learning support;* Step 1.2. Select the awareness dimensions. Identify the related awareness dimensions that will be considered in the assessment. – Awareness dimensions: *full awareness assessment (considering the workspace, collaboration, and contextual perspectives);* Step 1.3 Select the goals to be measured. We select which design categories are relevant in the collaborative environment for each awareness dimension.

continues on the next page

Table 18: Planning protocol (Moodle scenario) (continuation)

	– Goals: *to access all awareness dimensions;*
Phase 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. – Features: *all common Moodle features (assessment of the basics or most common Moodle resources, like chat, forum, assignments, lesson, and wiki (considering the basic functionalities available for interaction)* ; Step 2.2 Define participants and tasks. Identify the participants involved and the tasks that must be carried out within the collaborative environment. – Participants: *participants are invited voluntarily from the general public (undergraduate, postgraduate students and educators who use Moodle environments in their academic activities);* – Tasks description: *each participant answers the questions considering their Moodle perspective (it is not necessary to perform any procedure in the environment to answer the questionnaire);* Step 2.3 Identifies the Boundary, Persona, and Historical implications. The awareness information can be categorized in the boundary (physical, virtual, or both), persona (individual, other participants, group as a whole, groupware/system), and historical awareness perspectives (past, present, future). –Implications: *boundary (both – physical and virtual perspective); persona (individual, other participants, and group perspective); historical awareness (present – simultaneous interaction)*;
Phase 3. Planning the intervention	Step 3.1. Select the additional factors to access (like demographic, usability, and user experience). – Additional factors: *demographic analysis (age, gender, familiarity with the evaluated system, and related expertise in groupware and awareness concepts)*; Step 3.2. Generate the data collection instrument and prepare/customize the data collection instrument, considering the raised in activities 1.2, 1.3, and 2.3. – Data collection: *Assuming the BIB strategy* (HINKELMANN; KEMPTHORNE, 2005) *, we compiled the blocks (treatments) into ten different test books. Then, we set up a printed questionnaire and an online version (Google Forms) to collect participant feedback.*

8.2 PHASE 2: EXECUTION

The assessment items were prepared as described in Section 5.6, Tables 12 to 11. The assessment questionnaires (printed version) are available in Appendix C, Figures 68, and 70 to 79. Case study materials and the IRT dataset (Moodle scenario) are available at Mantau and Benitti (2023).

8.2.1 Model calibration

After applying the questionnaires, all observations were compiled into a .csv file. To calibrate our model, we ran the IRT script available at the assessment model repository (MANTAU; BENITTI, 2023) 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 16).

We analyzed the observed frequencies of each response category for all questionnaire items and grouped those with a small number of responses (HAIR et al., 2009) ($<$ 10 observations for each category). In these cases, we combined the response categories “strongly disagree” with “disagree”, or even the categories “agree” with “strongly agree”. 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 12 in Section 5.6), Q1 - goal, Q25 - feedback, 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 this item, the observed frequencies indicate that almost all participants could identify this information and mostly assign the category “agree” or “strongly agree”. From the collaboration awareness perspective (see Table 13 in Section 5.6), we disregard the assessment item Q33 - identity in the calibrated model. Participants generally indicated ease in identifying this assessment item and chose answers agreeing with the statement. From the contextual awareness perspective (see Table 14 in Section 5.6), Q68 - user mobility, Q70 - voice cues, and Q73 - map views, were removed from the results, indicating that these resources were absent or had not been used by the participants to collaborate.

Tables 19 to 21 present the coefficients of discrimination (a) and difficulty (b), the observed frequencies and Cronbach’s alpha coefficient (α) for the awareness taxonomy items. Appendix D, Figures 80 to 82 summarizes the items’ information functions for each awareness dimensions. The coefficients $b1$, $b2$, and $b3$ are related to each response category. Thus, for the items on the 4-point gradual scale, $b1$ represents the 1st category; $b2$ represents the 2nd category; $b3$ represents the 3rd category; the complement represents the 4th category.

When estimating the parameters, several items were observed with less than ten respondents in the category, meaning they needed to be grouped into the neighboring category. In this case, it is necessary to collapse adjacent categories, which are not reflective of unique positions along the latent trait scale, due keeping them as standalone categories could provide misleading information (COLVIN; GORGUN, 2020). For this reason, the grouping strategy is necessary to estimate the IRT item parameters well because the gradual scale can be difficult to estimate accurately without an adequate number of respondents in each category (LINACRE, 1999; WIRTH; EDWARDS, 2007).

For the items where grouping was applied, we used the 3-point gradual scale; therefore, only the parameters of $b1$ and $b2$ were generated. The NA values represent the cases where grouping was necessary.

Table 19 – Workspace awareness coefficients and observed frequencies (Moodle scenario)

Our Awareness Taxonomy		Item	Items' coefficients (a and b)				Observ. frequencies gradual scale (1 to 4)				Alpha (α)
			a	$b1$	$b2$	$b3$	**1**	**2**	**3**	**4**	
Involves activities	Subject	Q2	0,7932	-1,5414	0,9193	NA	NA	16	26	22	0,7090
	Content	Q3	1,5858	-2,1304	-0,2006	NA	NA	10	18	30	0,7122
	Motivation	Q4	0,6778	-2,0706	2,4024	NA	NA	14	40	12	0,7156
	Time required	Q5	1,1163	-0.9699	0,8077	NA	NA	16	22	18	0,7138
	Progress level	Q6	1,1191	-1,4770	-0,5139	1,2295	11	14	24	15	0,7130
	Help needed	Q7	0,8095	-2,2268	-0,0717	1,9644	11	21	21	13	0,7098
	Evaluation	Q8	1,0436	-0,5815	1,7182	NA	NA	25	28	11	0,7110
Consider workflow	Authorship	Q9	0,9934	-1,1884	-0,0252	NA	NA	16	13	31	0,7084
	Execution steps	Q10	2,6671	-0,5397	0,7748	NA	NA	17	26	13	0,7067
	Events and actions	Q11	1,5247	-1,3412	0,7280	NA	NA	12	34	20	0,7175
	Change location	Q12	1,0011	-1,0327	1,1827	NA	NA	15	23	14	0,7107
	Related activities	Q13	1,1207	-0,7287	1,1452	NA	NA	22	30	14	0,7072
	Parallel activities	Q14	1,9711	-0,3797	0,6649	NA	22	19	19	NA	0,7167
	Coordinated activities	Q15	1,4144	-1,4555	0,0240	1,5013	11	21	23	11	0,7189
	Adjusted activities	Q16	2,1575	-0,7385	1,0018	NA	NA	15	30	11	0,7153
Consider environment	Tools and materials	Q17	1,3923	-1,0215	0,8822	NA	NA	15	27	18	0,7123
	Artifacts and objects	Q18	1,7955	-0,4721	1,0942	NA	NA	20	25	11	0,7098
	Resources availability	Q19	1,8639	-0,8117	1,0347	NA	NA	13	28	11	0,7164

continues on the next page

Table 19: Workspace awareness coefficients and observed frequencies (Moodle scenario) (continuation)

	Critical elements	Q20	0,8361	-0,2750	1,9789	NA	NA	29	23	12	0,7135
	Virtual relationships	Q21	2,2106	-0,4563	0,7344	NA	NA	17	21	14	0,7029
Provide understanding	Meaning	Q22	1,3249	-1,2476	0,6147	NA	NA	12	25	19	0,7137
	Scenarios	Q23	1,6222	-0,3723	NA	NA	NA	20	32	NA	0,7088
	Sense-making	Q24	2,3510	-1,0805	-0,0625	NA	11	26	23	NA	0,7103
Allow interaction	Feedthrough	Q26	1,9898	-0,1167	NA	NA	NA	26	30	NA	0,7115
	Backchannel feedback	Q27	1,6779	-0,8088	0,6166	NA	14	21	17	NA	0,7019
	Feedforward	Q28	1,7109	-1,4260	-0,3490	1,2941	10	15	28	11	0,7215
Consider relationship	Action control	Q29	1,5130	-0,5363	0,7329	NA	18	18	16	NA	0,7094
	Access control	Q30	1,7946	-0,5250	1,4267	NA	NA	17	28	10	0,7164
	Access privileges	Q31	2,1936	-0,4831	NA	NA	NA	17	35	NA	0,7117
	Control mechanisms	Q32	1,2542	0,1271	NA	NA	NA	30	26	NA	0,7128

Table 20 – Collaboration awareness coefficients and observed frequencies (Moodle scenario)

Our Awareness Taxonomy		**Item**	**Items' coefficients (a and b)**				**Observ. frequencies** gradual scale (1 to 4)				**Alpha** (α)
			a	$b1$	$b2$	$b3$	**1**	**2**	**3**	**4**	
Allow identity	Shared profile	Q34	0,9156	-1,5332	1,1585	NA	NA	15	31	18	0,7049
	Preferences	Q35	1,7364	-0,6828	0,4634	NA	17	19	20	NA	0,7133
Consider capabilities	Rules	Q36	1,1751	-0,8591	-0,0157	NA	20	12	32	NA	0,7052
	Responsibilities	Q37	2,7275	-0,9733	0,0188	NA	11	18	27	NA	0,7056
	Privileges	Q38	1,6922	-0,7800	0,3323	NA	19	20	25	NA	0,7084
	Knowledge	Q39	0,8875	0,2306	NA	NA	NA	32	28	NA	0,7137
	Influences	Q40	1,8482	-1,1435	0,3305	NA	11	23	22	NA	0,7000
	Intentions	Q41	0,8823	0,5471	NA	NA	NA	34	22	NA	0,7106
Provide status	Availability	Q42	1,5714	-1,1277	0,3871	NA	13	22	25	NA	0,7161
	Presence	Q43	0,9444	-0,5023	1,8499	NA	NA	23	25	12	0,7167
	Activity level	Q44	1,9205	0,2047	NA	NA	NA	29	23	NA	0,7103
	Status	Q45	1,7488	-0,9669	0,2555	1,1967	15	22	17	12	0,7027
Provide communication	Mode	Q46	0,8922	-0,8545	1,5224	NA	NA	20	25	15	0,7130
	Connectivity	Q47	0,8982	-0,9201	1,1513	NA	NA	18	20	14	0,7104
	Message delivery	Q48	1,6645	-0,7523	0,7576	NA	NA	19	28	19	0,7095
	Message delays	Q49	0,7392	-1,3425	0,2116	1,9460	19	16	16	13	0,7004
	Interaction ways	Q50	1,0191	-1,8429	1,1364	NA	NA	11	37	18	0,7062
	Turn-talking	Q51	2,1391	-1,0444	0,0416	0,7849	12	17	14	17	0,7126
	Conversation	Q52	0,7549	-2,1610	0,9489	NA	NA	12	31	23	0,7050
Consider social	Expectations	Q53	1,4630	-0,8254	1,1236	NA	19	31	14	NA	0,7034
	Emotional status	Q54	1,2201	0,0808	1,5280	NA	34	20	12	NA	0,7139
	Non-verbal cues	Q55	1,0070	-0,6302	1,0609	NA	19	18	15	NA	0,7111

Table 21 – Contextual awareness coefficients and observed frequencies (Moodle scenario)

Our Awareness Taxonomy		Item	Items' coefficients (a and b)				Observ. frequencies gradual scale (1 to 4)				Alpha (α)
			a	$b1$	$b2$	$b3$	**1**	**2**	**3**	**4**	
Consider spatiality	Location	Q56	1,1738	0,1318	2,2303	NA	31	22	10	NA	0,7159
	Distance	Q57	1,6981	0,1254	2,1080	NA	28	21	10	NA	0,7166
	Restrictions	Q58	1,9528	-0,5234	1,2491	NA	17	27	11	NA	0,7175
	Places	Q59	1,5509	-0,2002	NA	NA	NA	25	31	NA	0,7091
	Topology	Q60	1,2331	-1,5816	0,1171	NA	11	23	30	NA	0,7057
	Attributes	Q61	1,5740	-1,1209	0,2966	NA	11	19	22	NA	0,7107
	View	Q62	2,0733	-1,0688	-0,0563	NA	11	16	29	NA	0,7121
	Reach	Q63	1,3554	-0,7246	1,2164	NA	20	31	15	NA	0,7006
	Orientation	Q64	1,1617	-1,1339	1,1300	NA	17	31	16	NA	0,7065
	Movement	Q65	1,1767	-1,2203	1,0664	NA	14	29	17	NA	0,7041
	Range of attention	Q66	1,6588	-1,2630	0,4149	NA	12	28	24	NA	0,7091
Allow mobility	User modality	Q67	0,7105	-2,0574	0,8696	NA	12	25	23	NA	0,7094
	Autonomy	Q69	1,2370	-1,3935	0,8088	NA	12	28	20	NA	0,7095
Provide navigation	Portholes/peepholes	Q71	1,8241	-1,0505	-0,0142	NA	12	16	29	NA	0,7120
	Eye-gaze cues	Q72	1,5464	-0,0853	1,6013	NA	31	26	10	NA	0,7080
	Viewports/Teleports	Q74	1,1298	-0,6132	1,5116	NA	20	25	11	NA	0,7055
	Objects location	Q75	1,4786	-0,7082	NA	NA	NA	16	36	NA	0,7130

8.2.2 Results

We obtained the voluntary participation of 149 students (78 in the online version and 71 in the printed one). As demographic data, we collected age, gender, expertise in 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 red dotted line – interval $[-2.35, +2.35]$). Table 22 presents the demographic distribution of the mean ability scores obtained (mean θ) for each demographic facet evaluated.

Regarding gender, we collected 104 male observations (70%) and 39 female observations (26%); 6 participants did not answer this question or chose the "other gender" option (4%). We collected 123 observations from individuals aged 18 to 28 years (83%), 22 from individuals aged 29 to 39 years (15%), and four from individuals between 40 and 50 years old (2%). No one under the age of 18 or over 50 years old participated in this research.

Table 22 – Demographic distribution (global scale)

(a) Age

Obs.	Group	mean θ
123	18 to 28 years	-0.0764
22	29 to 39 years	0.2818
4	40 to 50 years	0.4750

(b) Gender

Obs.	Group	mean θ
104	Male	-0.0038
39	Female	0.1051
6	Other	-0.8333

(c) Familiarity (awareness)

Obs.	Group	mean θ
75	1 - Novice	-0.2543
169	2 - competent	-0.0550
141	3 - proficient	0.0520
37	4 - expert	0.7692

(d) Familiarity (collaboration)

Obs.	Group	mean θ
46	1 - Novice	0.0353
40	2 - competent	-0.2273
50	3 - proficient	0.0343
13	4 - expert	0.4952

(e) Familiarity (environment)

Obs.	Group	mean θ
6	1 - Novice	0.0353
34	2 - competent	-0.2273
71	3 - proficient	0.0343
38	4 - expert	0.4952

We analyzed the normal distribution and mean score grouped by each demographic perspective to verify whether the model presents a distinction in different groups' discrimination values (or average). As shown in Figures 24 to 28, the normal curves generated for each group were significantly close, indicating that our model did not present considerably distinct behaviors in the observed groups and the cumulative probability distribution. As we can see, the sigmoid function suggests that the model does not significantly differentiate the IRT parameters of discrimination (a – the sigmoid slope) and difficulty (b – the sigmoid midpoint).

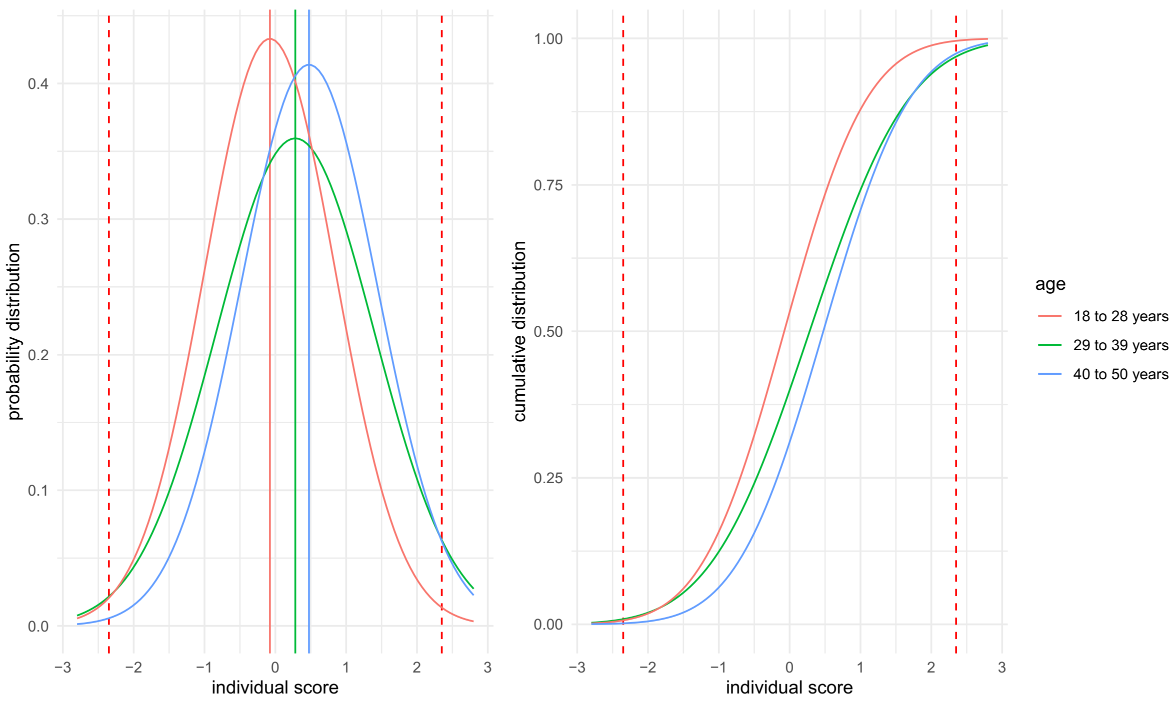
In the demographic facet of participants' age (Figure 24), the group of young individuals, 18 to 29 years old, gave a slight left-shift in the sigmoid function, demonstrating that, in general, older people have more straightforward use of these environments compared to younger. This factor may have positively corroborated the score because the sample of individuals in the first age group was significantly larger, and the obtained score of the older groups was influenced by the small variability of the collected sample (few observations). We did not obtain a sample of individuals under 18 or over 50 years, so constructing the scale for these age groups was impossible.

Grouping the participants by gender (Figure 25), we demonstrated that the model does not present additional difficulties or differentiate male or female participants. As can be seen, the distribution of individual scores for both groups was very close. Despite the sample mainly being composed of males and females, we observed that participants assigned to other gender options (4% of the total observations) achieved slightly lower scores when comparing the sigmoid functions. Due to the small sample, conducting a more detailed analysis of the instrument's behavior on this participants group is incipient.

By analyzing the participants' individual skill histograms (Figures 26a, 27a, and 28a), namely, familiarity with the e-learning supporting environment, collaboration, or awareness concepts, both normal distribution and probability cumulative distribution (Figures 26b, 27b, and 28b) were compatible with the participant's judgment.

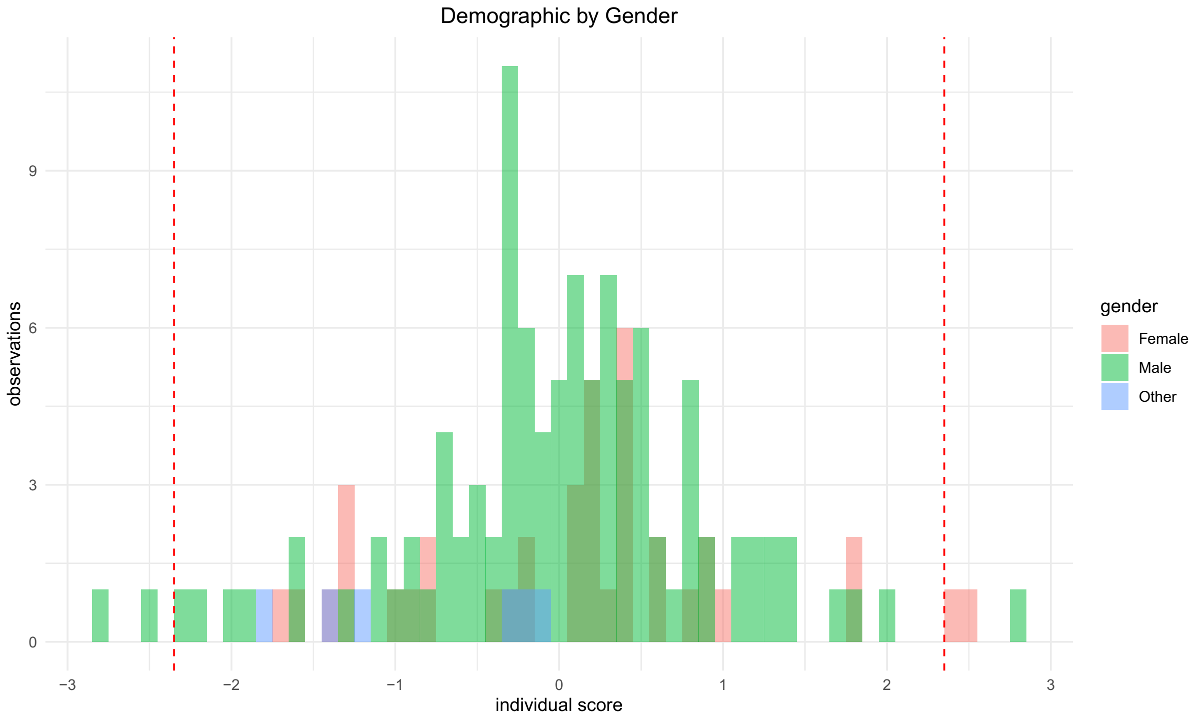
The observed frequencies in the histograms indicate a normal distribution for all groups and encompass the entire spectrum of the ability scale. Regarding familiarity with the evaluated environment, collaboration, and awareness concepts, novice participants presented a slightly distorted ability function concerning the other groups. This indicates that the evaluation model captured the relationship between the score obtained and the participant's ability. Comparing the demographic distribution of individual scores, we evidence that the greater the participant's experience, the better the score.

(a) Score

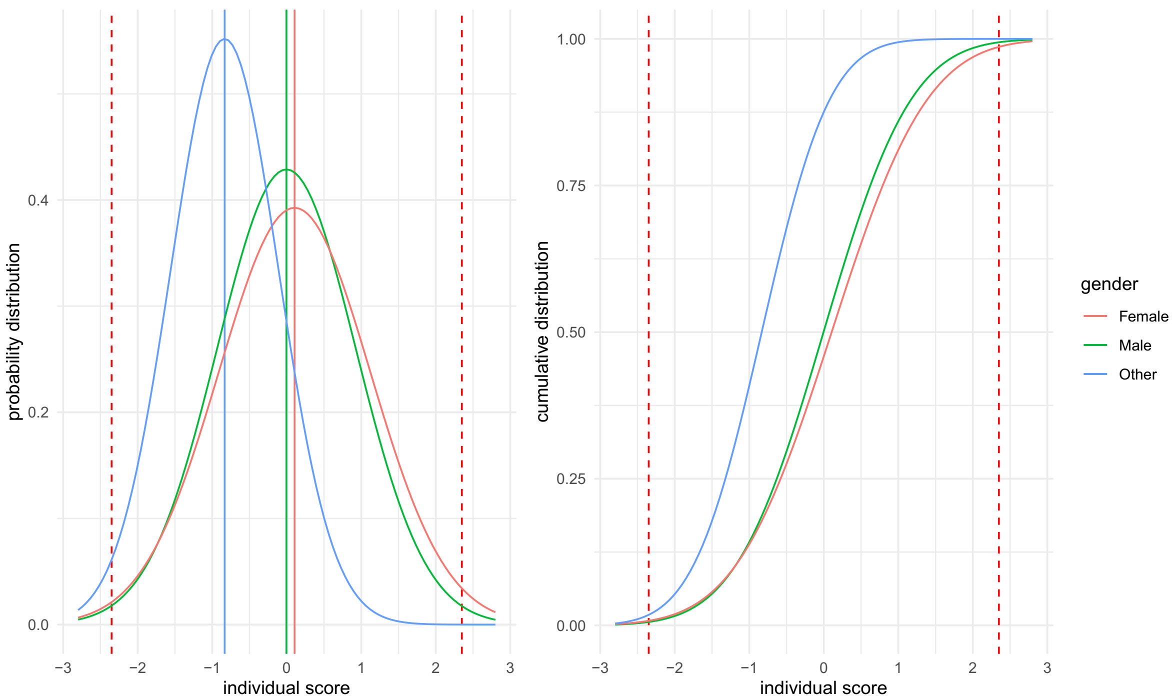


(b) Normal curves

Figure 24 – Demographic distribution of individual score by age (Moodle scenario)

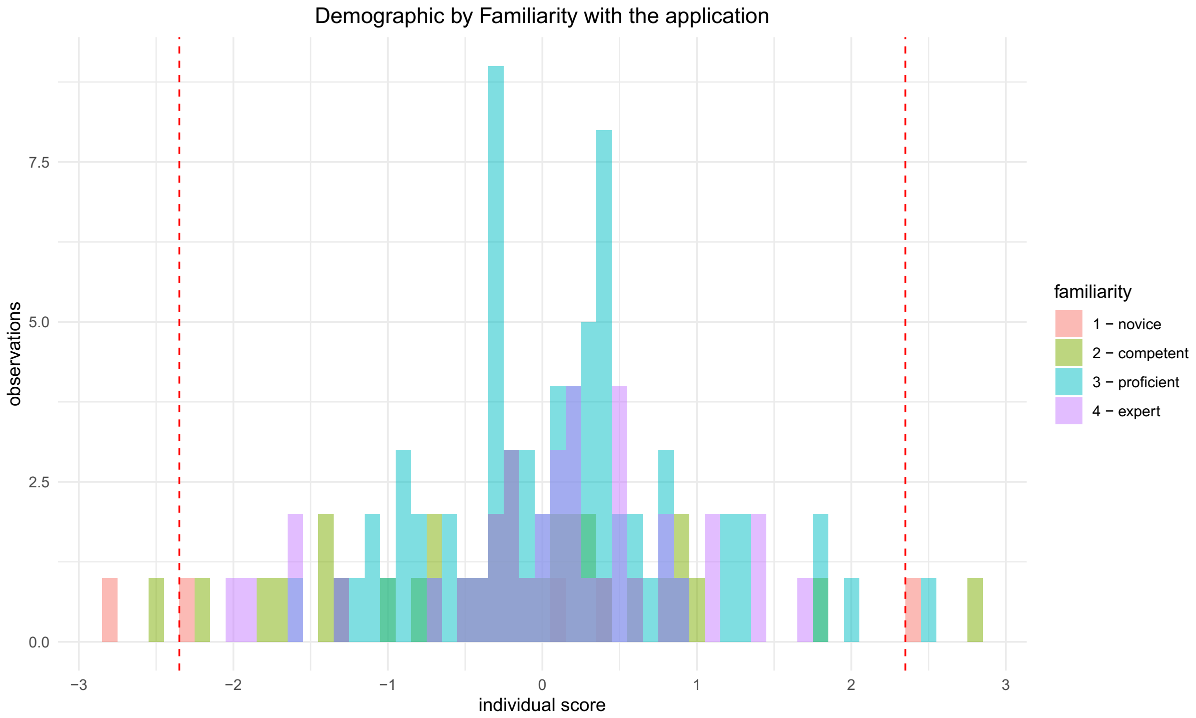


(a) Score

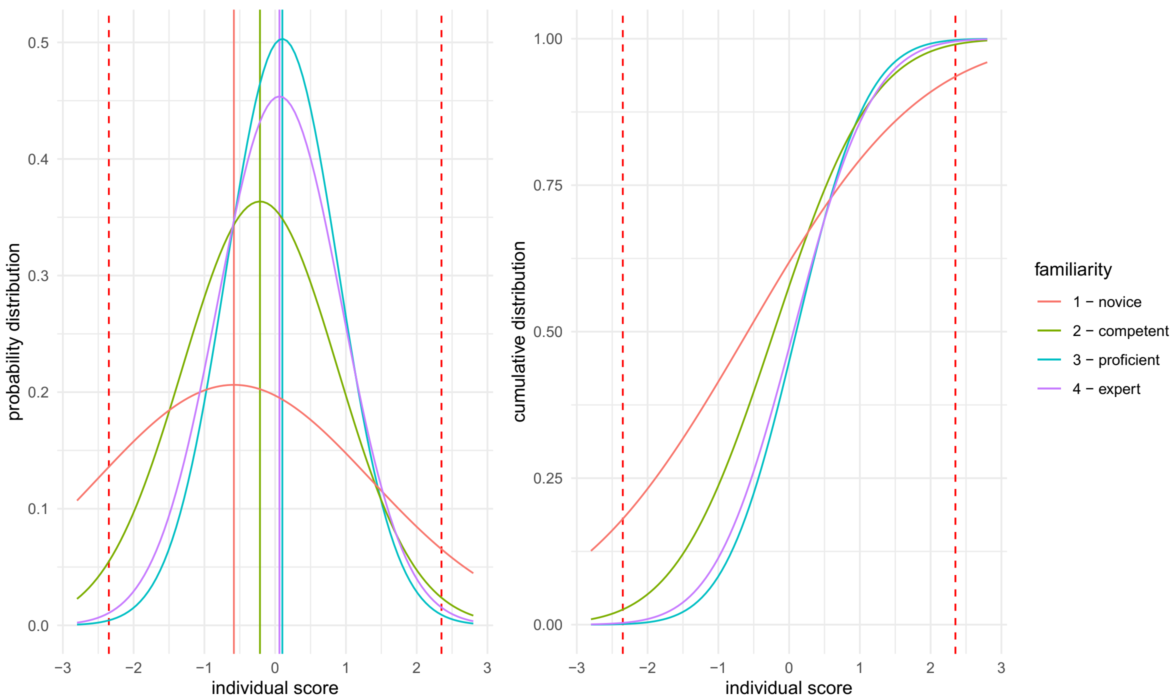


(b) Normal curves

Figure 25 – Demographic distribution of individual score by gender (Moodle scenario)

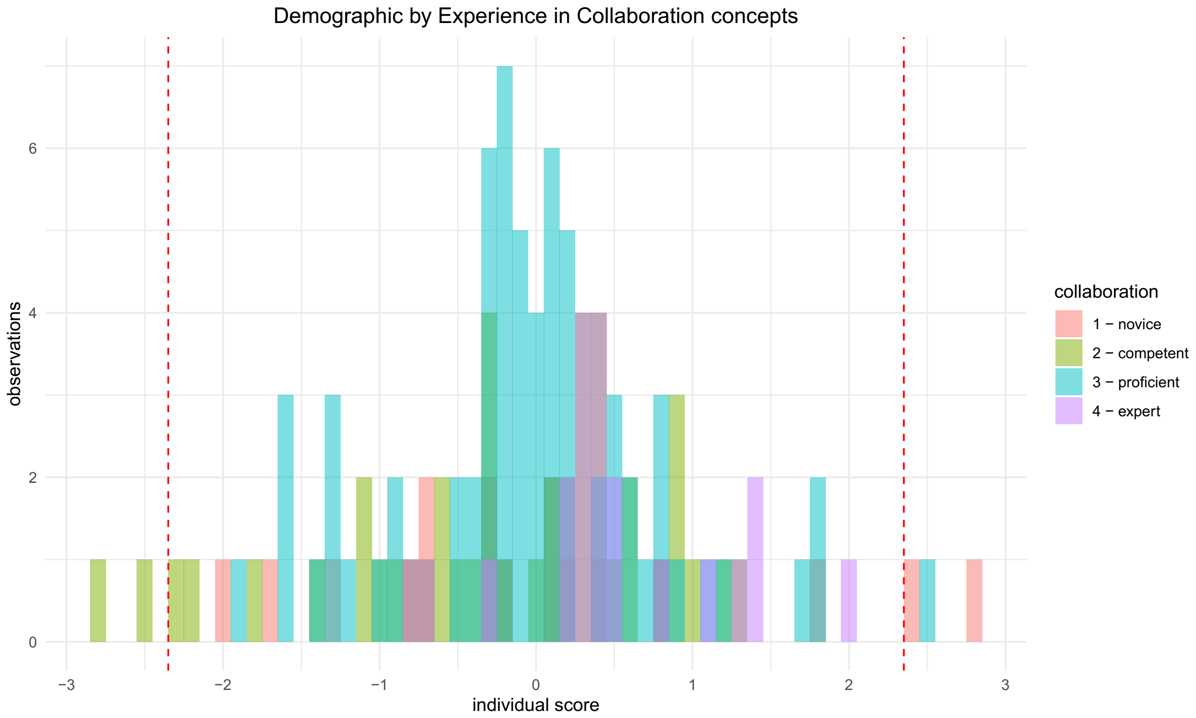


(a) Score

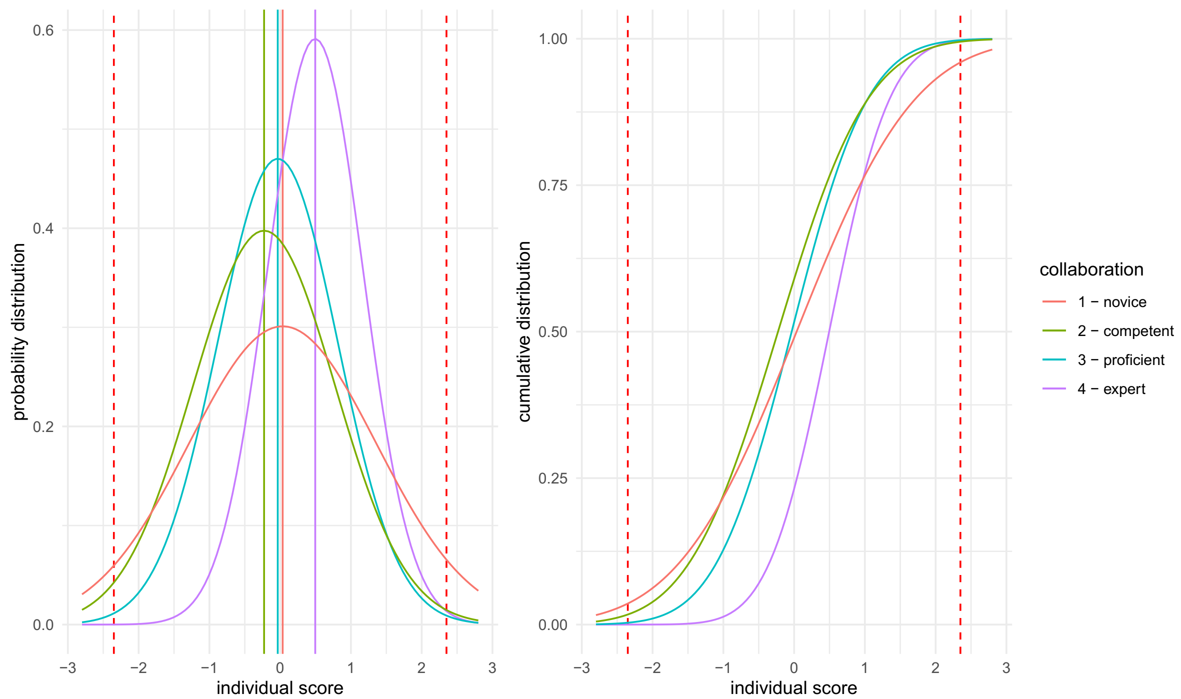


(b) Normal curvesy

Figure 26 – Demographic distribution of individual score by familiarity (Moodle scenario)

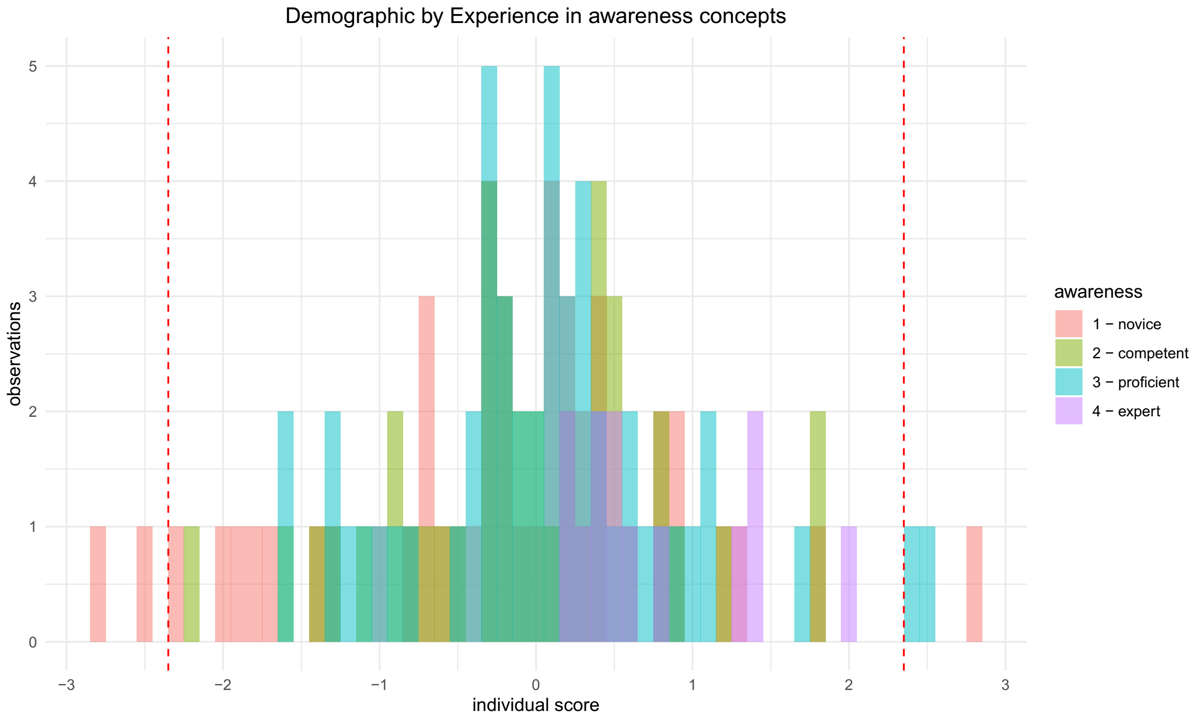


(a) Score

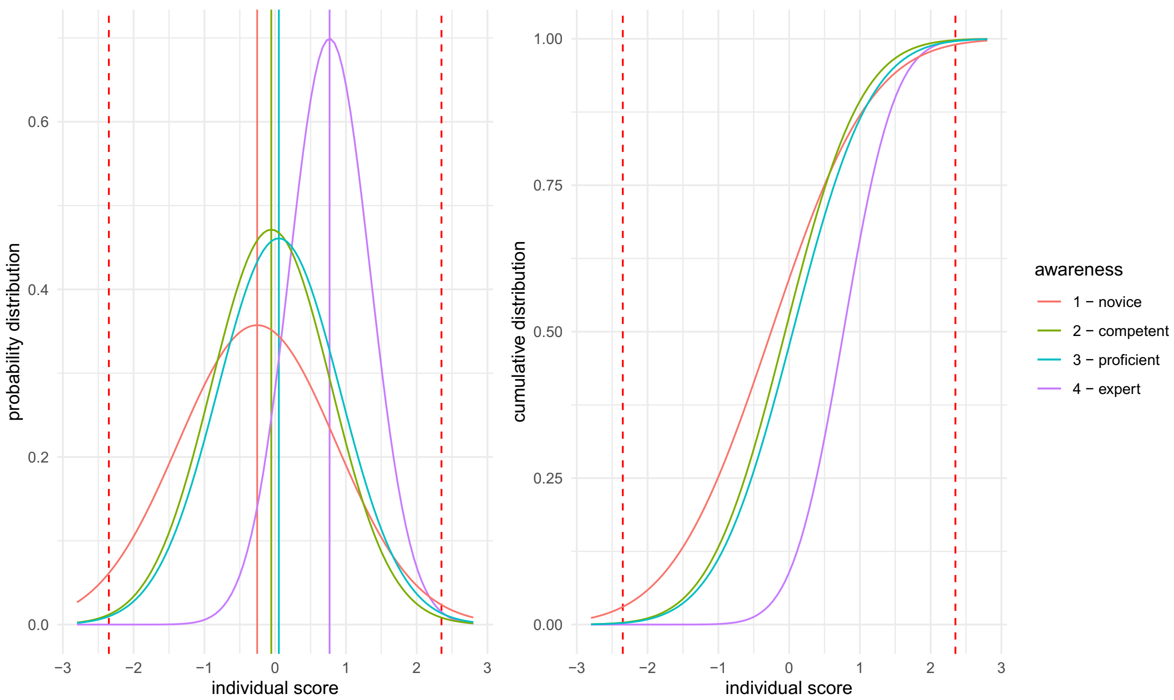


(b) Normal curves

Figure 27 – Demographic distribution of individual score by collaboration (Moodle scenario)



(a) Score



(b) Normal curves

Figure 28 – Demographic distribution of individual score by awareness (Moodle scenario)

For each awareness mechanism of our taxonomy (described in Section 4), 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 demonstrated.

Figure 29 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 (THISSEN; WAINER, 2001); the red dotted line represents the standard error $SE(\theta)$. The intersection point represents the limits at which the model is more representative.

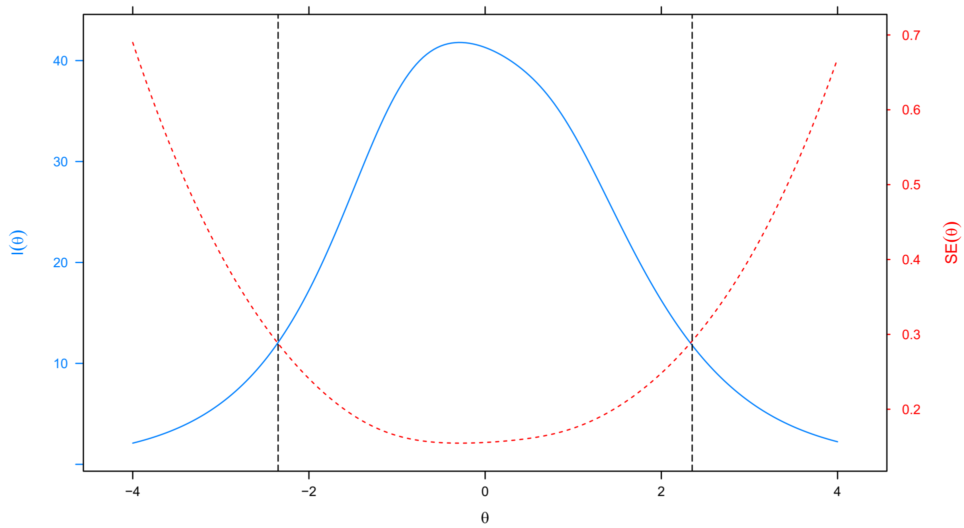


Figure 29 – Test information and SE (Moodle scenario)

This graph represents the region of the ability scale θ_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 $[-2.35, +2.35]$.

8.2.3 Assessing the awareness support scale

Applying the awareness measurement formulas 15 to 17 defined in Section 5.7.1, we calculated the probability scales $P_{i,k}(\theta_j)$ for the assessment element through our IRT awareness assessment model.

To construct our awareness support scale, we assume a coverage interval of $[-4.0, +4.0]$, although our model in this scenario is representative at the interval $[-2.35, +2.35]$. The scale $[-4.0, +4.0]$ is commonly used in the IRT models (DEVELLIS, 2016). Figure 30 presents the probability scales generated for each assessment item and awareness dimension.

Positioning the assessment items throughout our awareness scale, we defined four levels of quality: no quality ($\theta < -1$), low quality ($-1 \leq \theta \leq 0$), good quality ($0 \leq \theta \leq 1$), and excellent quality ($\theta > 1$). Tables 23, 24, and 25 exemplify the collaborative environment classification through our awareness quality scales. To position the respondents' ability score over the awareness scale, we use the theta parameter (θ), representing each subject's competence score.

In our 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 workspace, collaboration, and contextual awareness mechanisms.

In the ability level scales perspective, as shown in Figure 30, we have access to the general performance of the evaluated environments by each assessment item and awareness perspective. In our awareness scale, we represent the expected ability intervals in which participants present a certain probability P_i of selecting each response category presented, that is, the probable intervals $P_i(\theta)$ that participants are most likely to correctly identify/understand the awareness mechanism in the evaluated interface.

In the workspace, collaboration, and contextual perspective presented in Tables 23, 24, and 25, we categorized the results concerning the skill levels of the expected participants. Then, we established three participant ability intervals, describing the expected competencies concerning the awareness mechanisms participants in each ability score interval understand. Score values $-2.35 \leq \theta < -1$ indicate no quality; $\theta \leq -2.35$ or $\theta \geq 2.35$ represent model outliers.

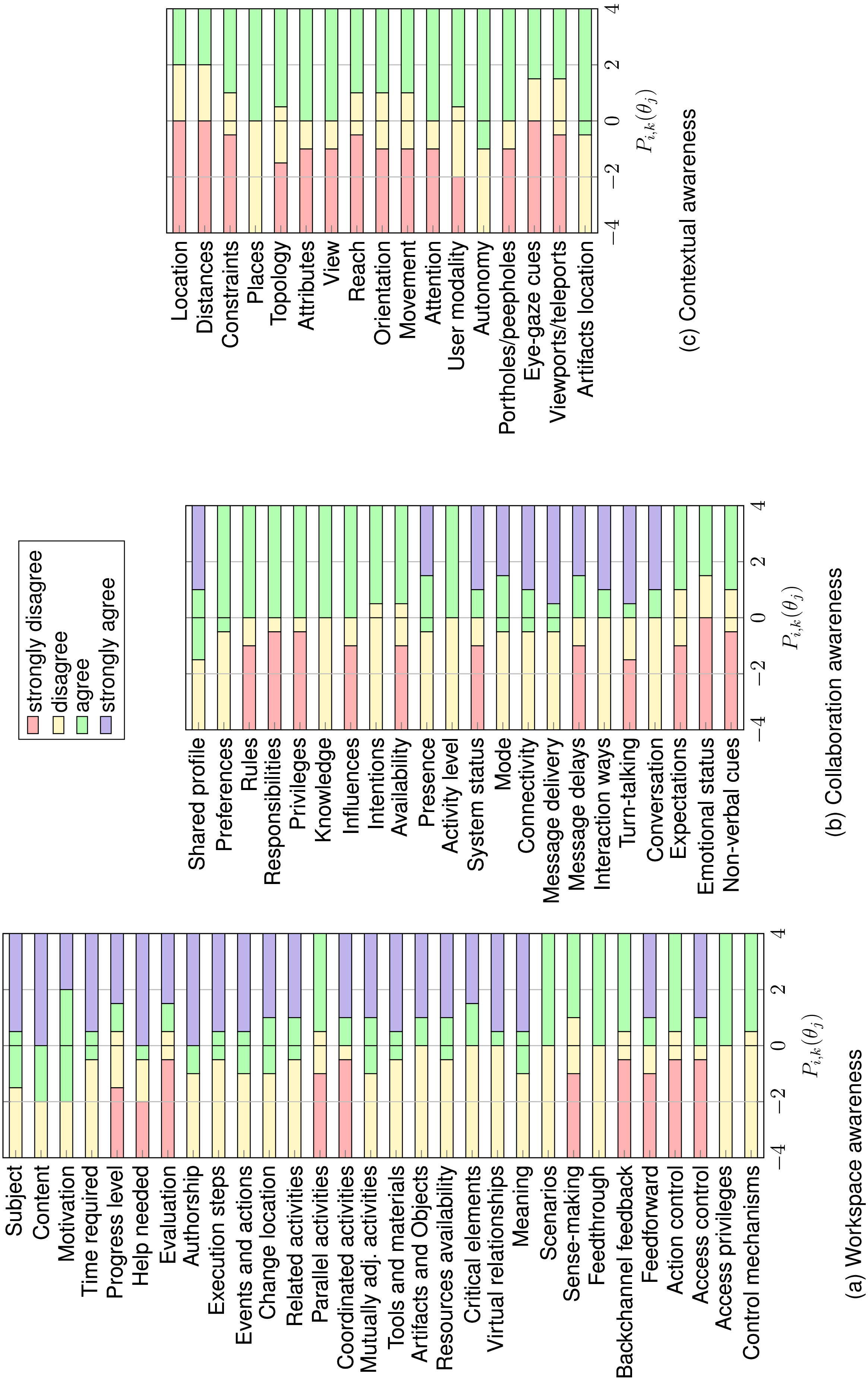


Figure 30 – Ability level scales (Moodle scenario)

Table 23 – Workspace awareness scales (Moodle scenario)

Level	Quality description
Low $-1 \leq \theta \leq 0$	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 $(0 \leq \theta \leq 1)$	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 they usually do not present an attractive design or good operability over the provided awareness mechanisms.
Excellent $(\theta > 1)$	At this level, the collaborative environment provides basic support for workspace capabilities and provides some information about the activities, environment, and workflow. The environment provides interaction and understanding of artifacts and objects shared in the workspace. In terms of usability, the workspace elements present considerable operability.

Table 24 – Collaboration awareness scales (Moodle scenario)

Level	Quality description
Low $-1 \leq \theta \leq 0$	The collaborative environment rarely provides social interaction and collaboration aspects. The environment hardly provides status and identity information or considers the participant's capabilities. Due to these limitations, the collaboration aspects are limited or even absent.
Good $(0 \leq \theta \leq 1)$	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 in the collaboration process.
Excellent $(\theta > 1)$	At this level, the collaborative environment is challenging for group members and presents few difficulties for interaction. It relates to participants' interests and provides ways for social interaction. In terms of operability, it has clear rules and is easy to interact with others.

Table 25 – Contextual awareness scales (Moodle scenario)

Level	Quality description
Low $-1 \leq \theta \leq 0$	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 $(0 \leq \theta \leq 1)$	The collaborative environment provides small access to contextual information. Participants partially reach environmental navigation and spatiality aspects. The environment provides some operability over participants' contextual information. Group members have difficulty identifying and appropriating the contextual information, and the contextual support in the environment remains questionable.
Excellent $(\theta > 1)$	At this level, the collaborative environment provides access to contextual information, and the environment presents some operability over participants' contextual information.

8.3 PHASE 3: REFLECTION

In this scenario, we evaluate the global awareness support provided by the Moodle environment, particularly about basic interaction functionalities such as chat, forum, assignments, lessons, and wiki. Using the awareness support scale, we identified three gradual awareness levels related to each expected participant's skill level.

As a first result, the scale allows us to point out the ability level a given participant must have to acquire the awareness mechanism in the collaborative environment properly. In this way, by observing the individual scale from an individual ability perspective (Figure 30), we can identify if a determined or even a set of awareness elements are accessed and what effort is involved.

In other words, the results can be transcribed concerning the elements that are or are not perceived at each ability scale interval – as exemplified in the workspace, collaboration, and contextual awareness scales (Tables 23, 24, and 25). Furthermore, by establishing three different levels of participant expected ability, it is possible to identify if each awareness mechanism is captured or not and, considering the degree of agreement to the presence of this resource (strongly disagree to strongly agree), we can measure the level of support is necessary to each participant.

8.3.1 Scale interpretation

From the workspace awareness perspective, hardly awareness support mechanisms are received for participants with small individual ability $(\theta < -1)$[1]. In short, the higher the participant's degree of ability, the better the score obtained in the assessment of environmental awareness support. This also applies to collaboration awareness and is even more accentuated in the contextual awareness perspective.

Progress level, help needed, evaluation, parallel activities, coordinated activities, sense-making, backchannel feedback, feedforward, action control, and access control awareness mechanisms demanded a greater ability to be recognized in the environment (with θ close to 0). Additionally, in parallel activities, sense-making, backchannel feedback, and action control mechanisms, this greater demand was observed across the entire scale (none or few participants strongly agree with the existence of these elements).

Scenarios, feedthrough, access privileges, and control mechanisms awareness elements also showed greater difficulty for participants to agree with these resources strongly. On the other hand, awareness cues such as subject, content, motivation, authorship, events and action, change location, mutually adjusted activities, resources availability, and meaning were well evaluated by participants with lower skill levels (novices). This strongly suggests that these mechanisms are adequately supplied by the collaborative tool.

From the collaboration awareness perspective, rules, responsibilities, privileges, influences, availability, system status, message delays, turn-talking, expectations, emotional status, and non-verbal cues demanded a greater ability to be recognized in the environment, and participants with less ability $(\theta < 0)$ generally strongly disagreed with the appropriation of these resources. Of these, only system status, message delays, and turn-talking were satisfactorily absorbed by more skilled participants, suggesting improvements. Participants pointed out a better appropriation of awareness resources at lower skill levels (novices) regarding shared profile, presence, mode, connectivity, message delivery, interaction ways, and conversation. System status and message delays were also well evaluated in higher ability scores.

In preferences, rules, responsibilities, privileges, knowledge, influences, intentions, availability, activity level, expectations, emotional status, and non-verbal cues, participants were unlikely to strongly agree with the presence of these elements – which suggests that there are difficulties or noise in their understanding; as a recommendation, a thorough analysis of collaboration resources should be considered.

[1] This represents novice participants below one standard deviation from the observed mean $(\theta = 0)$.

From the contextual awareness perspective, we can identify a marked difficulty in understanding the awareness mechanisms evaluated. Logically, we have to consider whether these mechanisms are less relevant to the target context or are not this environment's key purpose.

Although a more in-depth analysis of the missing contextual aspects is not encouraged, some interesting points can be observed. First, only three support mechanisms did not obtain a strongly discordant positioning for participants with lower ability scores (θ closer to 0): places, autonomy, and artifacts location. Second, in none of the mechanisms evaluated was the result suggestive that, at higher skill levels, participants strongly tend to agree with the presence of these resources. In conclusion, from a contextual point of view, the evaluated environment presents a great lack of support that must be considered.

8.4 MODEL RELIABILITY

The reliability of a set of items is one of the properties to evaluate the quality of the instrument. One of the ways to check internal consistency is through Cronbach's alpha coefficient (α) (DEVELLIS, 2016). 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 (θ)'s scale scores. We use the IRT technical properties of discrimination (a) and difficulty (b), combined with α coefficient, to assess the Awareness Assessment Model Instrument reliability and internal consistency.

In this scenario, both alpha and IRT parameters corroborate the validation of our proposed model. First, the adequate representation of the awareness scale θ (interval $[-2.35, +2.35]$ as presented in Figure 29) is good evidence of the instrument's reliability. In addition, the internal reliability through Cronbach's alpha coefficient (DEVELLIS, 2016) demonstrated an acceptable internal consistency for all assessment items $(\alpha > 0.70)$. Second, we calculate the instrument's reliability function $r_{xx}(\theta)$ (THISSEN; WAINER, 2001; CHALMERS, 2012), across participants' latent trail (see Figure 31). Our model shows excellent reliability[2], and the function reaches its highest value $(r_{xx} > 0.90)$ over the scale region where the information function is representative.

[2] 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 (DEVELLIS, 2016).

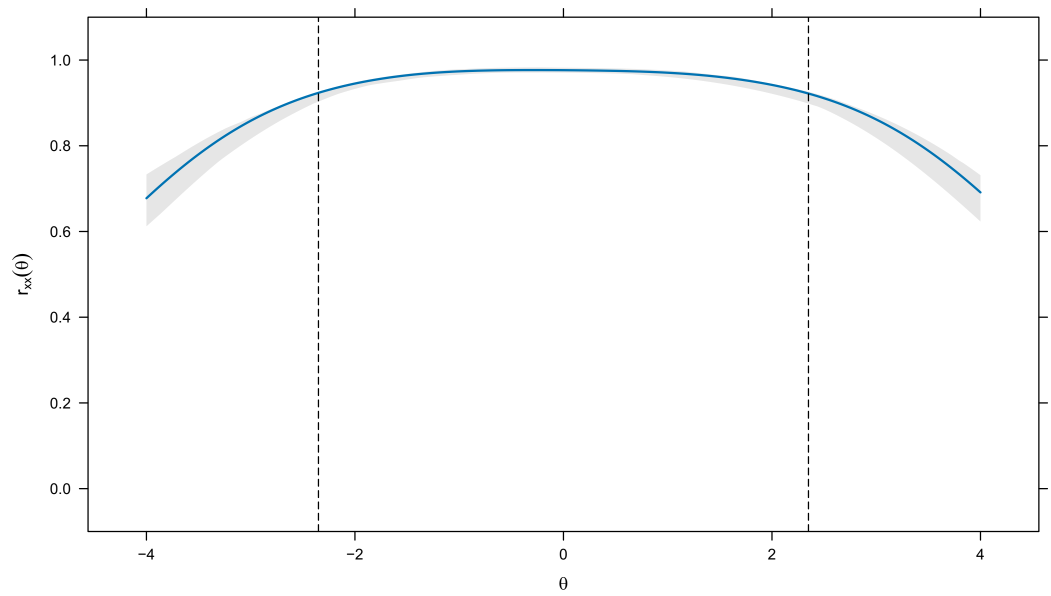


Figure 31 – Test reliability (Moodle scenario)

8.5 DISCUSSION

In this scenario, we obtained voluntary participation from 149 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. The skill and awareness quality scales were then constructed based on the 69 calibrated items.

The results of the Moolde assessment were positive. 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 lower ability scores hardly identify existing awareness mechanisms.

Estimating IRT parameters with a low standard error and positioning items on the scale requires many respondents per item category. We consider this a limiting factor for better calibration and adjustment of the awareness scale in this specific scenario. Intending to calibrate the model, we adopted the grouping technique (LINACRE, 1999; WIRTH; EDWARDS, 2007); thus, we adjusted the response categories of many assessment items, converting from a gradual 4-point scale (strongly disagree to strongly agree) into a three-point scale or even a dichotomous perspective.

In cases where the grouping was applied, either for the negative response (disagree or strongly disagree) or positive ones (agree or strongly agree), the degree of agreement with the response item is lost; the model only captures whether the participant presents a negative tendency (disagrees) or positive (agrees) with the evaluation statement presented. It is necessary to carry out more comprehensive scenarios to verify the model behavior and calibrate the assessment items for larger samples.

In some cases, the non-calibration occurred due to many positive (agree or strongly agree) or negative (disagree or strongly disagree) responses. In the first, our 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.

8.5.1 Observed limitations

We identified three basic limitations when carrying out this scenario:

First, the small number of respondents (149) is the main limitation in this scenario. 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.

Due to the limited number of observations collected, grouping some answer categories in most evaluation items was necessary. Consequently, the model converged to a representative scale between intervals [-2.35, +2.35]. On the other hand, by observing the positioning of items on the scale, it is possible to realize that the model starts to be more representative of $\theta > -1$ (intervals [-1, +2.35]). Although demonstrating the construction of the awareness support quality scale is possible through IRT, new investigation scenarios are needed to calibrate the awareness scale.

Second, the IRT model could not calibrate some items from the 75 assessment items initially proposed in our taxonomy, as discussed in Section 8.2.1. In addition, due to the small number of observations obtained, validating these items was also impossible in this case study; alternatively, we suggest new evaluation scenarios to investigate their suitability.

Third, the complexity and difficulty of performing IRT analyses. These analyses require specialized knowledge to perform tests of assumptions, estimation of parameters, and tests for model adjustment.

8.5.2 Related publications

The results presented in this chapter were published in:

- MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. The Awareness Assessment Model: measuring the awareness and collaboration support over the participant's perspective. In: Anais do XVIII Simpósio Brasileiro de Sistemas Colaborativos, SBSC. SBC, 2023. P. 30-43.

9 CASE STUDY 2: VIDEOCONFERENCING ENVIRONEMNTS

We live in a connected world where our daily social interactions are amplified through collaborative environments, whether through social, work, or teaching interactions. Our daily life is, respecting proportion, a mixture of face-to-face and virtual interactions. We believe that the impact/influence of an individual's interaction directly depends on its ratio (physical/virtual) and how satisfactory it is performed. Face-to-face interactions take advantage of this balance due to the inherently imposed limitations of the virtual environments (OULASVIRTA, 2008). The restrictions in virtual environments can arise from different sources, such as environmental and context restrictions, providing awareness and collaboration aspects, or even design issues (MANTAU; BERKENBROCK; BERKENBROCK, 2014).

In this scenario, we evaluated virtual collaboration environments intended for simultaneous communication/interaction between two or more people. For example, conference environments, videoconferencing, virtual events, webinars, etc. In these environments, 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.

9.1 PHASE 1: PLANNING

Initially, we planned the case study using the awareness assessment process (described in section 5.3). As a result of this step, we created the planning protocol artifact (Table 26), where we determined the intervention's objectives, scope, and life cycle.

Table 26 – Planning protocol (videoconferencing scenario)

Phase 1. Determine the objectives	Step 1.1. Define the assessment objectives. This activity defines the evaluation goal in terms of the object of study, purpose, perspective, and context. – Object of study: *videoconferencing environments;* – Purpose: *to evaluate virtual collaboration environments, designed for communication/interaction between two or more people simultaneously, for example, conference environments, video conferencing, virtual events, webinars, etc.*; – Perspective: *participant's perspective;* – Context or environment: *general purpose video conferencing environment, such as Google meeting, Microsoft Teams, Skype, Zoom, etc.*;

continues on the next page

Table 26: Planning protocol of videoconferencing scenario (continuation)

	Step 1.2. Select the awareness dimensions. Identify the related awareness dimensions that will be considered in the assessment. – Awareness dimensions: *full awareness assessment (considering workspace, collaboration, and contextual perspectives)*; Step 1.3 Select the goals to be measured. We select which design categories are relevant in the collaborative environment for each awareness dimension. – Goals: *to access all awareness dimensions;*
Phase 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. – Features: *all features (a general assessment of the videoconferencing environment considering the basic functionalities available for simultaneous interaction)*; Step 2.2 Define participants and tasks. Identify the participants involved and the tasks that must be performed within the collaborative environment. – Participants: *participants are invited voluntarily from the general public (undergraduate and postgraduate students who use videoconferencing environments in their academic activities);* – Tasks description: *each participant answers the questions considering the virtual collaboration environment (video conference) of their preference;* Step 2.3 Identifies the Boundary, Persona, and Historical implications. The awareness information can be categorized in the Boundary (physical, virtual, or both), Persona (individual, other participants, group as a whole, groupware/system), and Historical awareness perspectives (past, present, future). –Implications: *boundary (both – physical and virtual perspective); persona (individual, other participants, and group perspective); historical awareness (present – simultaneous interaction)*;
Phase 3. Planning the intervention	Step 3.1. Select the additional factors to access (like demographic, usability, and user experience). – Additional factors: *demographic analysis (age, gender, and related expertise)*; Step 3.2. Generate the data collection instrument and prepare/customize the data collection instrument, considering the raised in activities 1.2, 1.3, and 2.3.

continues on the next page

Table 26: Planning protocol of videoconferencing scenario (continuation)

– Data collection: *Assuming the BIB strategy* (HINKELMANN; KEMPTHORNE, 2005) *, we compiled the blocks (treatments) into ten different test books. Then, we set up a printed questionnaire and an online version (Google Forms) to collect participant feedback.*

9.2 PHASE 2: EXECUTION

The assessment items were prepared as described in Section 5.6 (Tables 12 to 11). The assessment questionnaires (printed version) are available in Appendix C, Figures 69 to 79. The case study materials and the IRT dataset are available at (MANTAU; BENITTI, 2023).

9.2.1 Model calibration

After applying the questionnaires, all observations were compiled into a .csv file. To calibrate our model, we ran the IRT script available in the assessment model repository (MANTAU; BENITTI, 2023) and interpreted the output values of discrimination (a) and difficulty (b) disregarding items with $a < 0.65$ or $a > 4.0$ (as defined in Section 5.6 Table 16).

We analyzed the observed frequencies of each response category for all questionnaire items and grouped those with a small number of responses (HAIR et al., 2009) ($<$ 10 observations for each category). In these cases, we combined the response categories “strongly disagree” with “disagree”, or even the categories “agree” with “strongly agree”. Then, we re-run the model with the remaining items and generated the final discrimination and difficulty coefficient.

The items of the workspace awareness assessment (see Section 5.6 Table 12), 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 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 Section 5.6 Table 13), 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 responses that agreed with the statement.

From the contextual awareness perspective (see Section 5.6 Table 14), the Mobility design category did not present any assessment elements that converged 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 (that is, map views, viewports/teleports, and artifact 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 27 to 29 present the coefficients of discrimination (a) and difficulty (b), the observed frequencies and Cronbach's alpha coefficient (α) for the awareness taxonomy items. Appendix E, Figures 83 to 85 summarizes the items' information functions for each awareness dimension.

The coefficients $b1$, $b2$, and $b3$ are related to each response category. Thus, for items on the 4-point gradual scale, $b1$ represents the first category; $b2$ represents the second category; $b3$ represents the third category; the complement represents the fourth category. For the items where grouping was applied, we used the 3-point gradual scale; therefore, only the parameters of $b1$ and $b2$ were generated. The NA values represent the cases where grouping was necessary.

Table 27 – Workspace awareness coefficients and observed frequencies (videoconferencing scenario)

Our Awareness Taxonomy		Item	Items' coefficients (a and b)				Observ. frequencies gradual scale (1 to 4)				Alpha (α)
			a	$b1$	$b2$	$b3$	**1**	**2**	**3**	**4**	
Involves activities	Motivation	Q4	1,1280	-0,9310	1,2833	NA	NA	53	79	33	0,9050
	Time required	Q5	1,0289	-2,5473	-0,5770	1,5185	14	44	71	42	0,9033
	Progress level	Q6	0,9899	-3,2187	-1,0421	1,0893	11	43	71	46	0,9043
	Help needed	Q7	1,5914	-2,0704	-0,3354	1,0043	13	59	60	33	0,9052
	Evaluation	Q8	0,8802	-3,6196	0,9669	1,6140	10	47	77	37	0,9044
Consider workflow	Authorship	Q9	0,6787	-3,5600	-0,7452	NA	NA	16	49	105	0,9057
	Execution steps	Q10	1,1596	-1,2761	0,8120	NA	NA	34	75	62	0,9031
	Events and actions	Q11	1,9203	-2,0476	-0,7914	0,6574	10	38	75	42	0,9048
	Change location	Q12	0,8123	-2,1632	-1,7750	NA	NA	29	78	60	0,9042
	Related activities	Q13	1,4849	-2,0482	-0,4084	1,5512	15	53	77	20	0,9052
	Parallel activities	Q14	1,2601	-0,8449	0,3192	1,9690	51	45	52	22	0,9041
	Coordinated activities	Q15	1,5772	-0,6339	-0,6721	NA	NA	57	62	46	0,9056
	Adjusted activities	Q16	1,1387	-1,0936	1,4960	NA	NA	40	91	40	0,9048
Consider environment	Tools and materials	Q17	1,0073	-1,0999	1,2883	NA	NA	48	78	44	0,9048
	Artifacts and objects	Q18	1,1366	-0,9178	1,2490	NA	NA	45	79	47	0,9040
	Resources availability	Q19	1,0884	-1,8381	0,6744	NA	NA	26	82	59	0,9043
	Critical elements	Q20	0,8541	-3,4585	-0,3474	1,9833	12	64	65	30	0,9044
	Virtual relationships	Q21	1,1404	-1,0545	1,2558	NA	NA	46	82	39	0,9049

continues on the next page

Table 27: Workspace awareness coefficients and observed frequencies (continuation)

Provide understanding	Meaning	Q22	0,7950	-2,3278	-1,2254	NA	NA	24	91	56	0,9041
	Scenarios	Q23	1,0154	-2,7104	-0,3739	1,4981	14	56	61	36	0,9060
	Sense-making	Q24	1,9497	-1,4203	-0,0625	1,1742	23	59	54	34	0,9029
Allow interaction	Feedback	Q25	1,1922	-0,9638	1,1656	NA	NA	42	71	52	0,9041
	Feedthrough	Q26	1,1922	-0,9638	1,1656	NA	NA	43	79	49	0,9054
	Backchannel feedback	Q27	1,1937	-1,6622	-0,0252	2,1606	28	55	66	18	0,9050
	Feedforward	Q28	1,1676	-2,2529	-1,0229	1,1216	19	32	79	41	0,9043
Consider relationship	Action control	Q29	1,1735	-1,7158	0,8844	2,7934	27	76	54	10	0,9046
	Access privileges	Q31	0,6591	-3,4114	-0,9647	1,4511	10	50	59	48	0,9064
	Control mechanisms	Q32	1,3180	-1,6571	-0,3972	1,1432	12	49	63	47	0,9036

Table 28 – Collaboration awareness coefficients and observed frequencies (videoconferencing scenario)

Our Awareness Taxonomy		**Item**	**Items' coefficients (a and b)**				**Observ. frequencies** gradual scale (1 to 4)				**Alpha** (α)
			a	$b1$	$b2$	$b3$	**1**	**2**	**3**	**4**	
Allow identity	Shared profile	Q34	1,1057	-2,1418	0,2212	NA	NA	23	75	73	0,9039
	Preferences	Q35	1,8723	-0,9100	0,5117	1,7782	35	69	46	21	0,9027
Consider capabilities	Rules	Q36	1,3296	-1,3154	0,3715	2,3652	39	67	53	12	0,9032
	Responsibilities	Q37	2,1670	-1,1895	-0,0012	1,3518	23	53	64	31	0,9026
	Privileges	Q38	1,1522	-1,6447	0,2793	2,0666	33	67	51	20	0,9046
	Knowledge	Q39	1,7850	-2,0176	-0,5474	1,0950	13	46	73	38	0,9034
	Influences	Q40	1,9900	-2,2568	-0,0188	1,5459	23	53	69	26	0,9030
	Intentions	Q41	1,8639	-1,8831	-0,1796	1,4568	11	56	74	30	0,9035
Provide status	Availability	Q42	1,7378	-1,4944	-0,2251	0,8393	25	50	41	54	0,9042
	Presence	Q43	0,9258	-2,4657	-1,1891	0,8851	20	26	52	72	0,9054
	Status	Q45	1,0102	-2,0529	-0,0633	0,7748	24	59	64	18	0,9043
Provide communication	Connectivity	Q47	0,8764	-1,9375	0,1782	NA	NA	31	59	77	0,9055
	Message delays	Q49	0,7535	-1,9543	-0,1323	1,8879	38	47	51	35	0,9076
	Interaction ways	Q50	1,2655	-1,8055	0,3198	NA	NA	23	78	64	0,9057
	Turn-talking	Q51	0,7284	-3,7535	-1,7365	-0,0993	13	29	41	87	0,9063
	Conversation	Q52	1,5022	-1,6849	0,1069	NA	NA	21	72	72	0,9054
Consider social	Expectations	Q53	1,2837	-1,3889	0,7965	NA	37	84	50	NA	0,9043
	Emotional status	Q54	1,6395	-0,5993	0,6637	NA	60	61	44	NA	0,9050
	Non-verbal cues	Q55	0,7994	-2,4854	-0,8492	1,1660	24	35	57	51	0,9055

Table 29 – Contextual awareness coefficients and observed frequencies (videoconferencing scenario)

Our Awareness Taxonomy		Item	Items' coefficients (a and b)				Observ. frequencies gradual scale (1 to 4)				Alpha (α)
			a	$b1$	$b2$	$b3$	**1**	**2**	**3**	**4**	
Consider spatiality	Location	Q56	0,8022	-0,5256	1,4806	3,8581	71	56	33	10	0,9048
	Distance	Q57	0,7796	-0,7070	2,3092	3,4549	64	76	14	13	0,9072
	Restrictions	Q58	1,1085	-1,2682	0,6947	2,3679	40	72	39	16	0,9056
	Places	Q59	1,0388	-1,9460	-0,0539	2,3235	23	55	71	22	0,9027
	Topology	Q60	1,0697	-2,2555	0,3576	2,2487	22	54	76	19	0,9047
	Attributes	Q61	1,2875	-2,1366	-0,1071	1,8837	16	65	65	21	0,9041
	View	Q62	0,7730	-2,5481	-0,0302	2,1880	22	50	66	33	0,9039
	Reach	Q63	0,9585	-1,6942	1,0402	2,8024	34	85	34	12	0,9031
	Orientation	Q64	1,2929	-1,5301	0,6473	2,3420	33	83	42	13	0,9029
	Movement	Q65	0,7958	-2,1675	0,2426	2,3270	29	62	51	28	0,9060
	Range of attention	Q66	1,5419	-1,4791	0,1489	2,1598	30	68	61	12	0,9036
Provide navigation	Voice cues	Q70	0,8120	-2,1272	0,6550	NA	NA	31	76	58	0,9055
	Eye-gaze cues	Q72	1,0567	-0,5244	0,9665	NA	68	53	44	NA	0,9046

9.2.2 Results

We obtained the 422 voluntary participation. As demographic data, we collected 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 – interval $[-2.96, +2.70]$). Table 30 presents the demographic distribution of the mean ability scores obtained (mean θ) for each demographic facet evaluated.

Table 30 – Demographic distribution (videoconferencing scenario)

(a) Age

Obs.	Group	mean θ
345	18 to 28 years	-0.0583
58	29 to 39 years	0.2603
17	40 to 50 years	0.1353
2	51 years or more	-0.5500

(b) Gender

Obs.	Group	mean θ
298	Male	-0.0188
112	Female	0.0205
12	Other	-0.0417

(c) Familiarity (awareness)

Obs.	Group	mean θ
75	1 - Novice	-0.1480
169	2 - competent	-0.0325
141	3 - proficient	0.0113
37	4 - expert	0.3027

(d) Familiarity (collaboration)

Obs.	Group	mean θ
26	1 - Novice	-0.1846
107	2 - competent	-0.0346
213	3 - proficient	0.0493
76	4 - expert	0.2000

(e) Familiarity (environment)

Obs.	Group	mean θ
15	1 - Novice	-0.5667
65	2 - competent	-0.3539
213	3 - proficient	0.0005
129	4 - expert	0.2140

(f) Environmmet

Obs.	Group	mean θ
39	Discord	0.2154
143	Google Meet	-0.1231
104	Microsoft Teams	0.1231
71	Moodle (BBB)	-0.2127
15	Skype	0.2000
35	Zoom	0.1600
15	Other	-0.0600

We obtained the voluntary participation of 422 participants. Regarding gender, we collected 298 male observations (70%) and 112 female observations (27%); 12 participants did not answer this question or chose the “other gender” option (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 also generated frequency histograms based on the participant's scores and demographic facets (Figures 32 to 37). To verify whether the model presents a distinction in discrimination values (or average) of different groups, we also calculated the normal distribution and the mean score grouped by each demographic perspective.

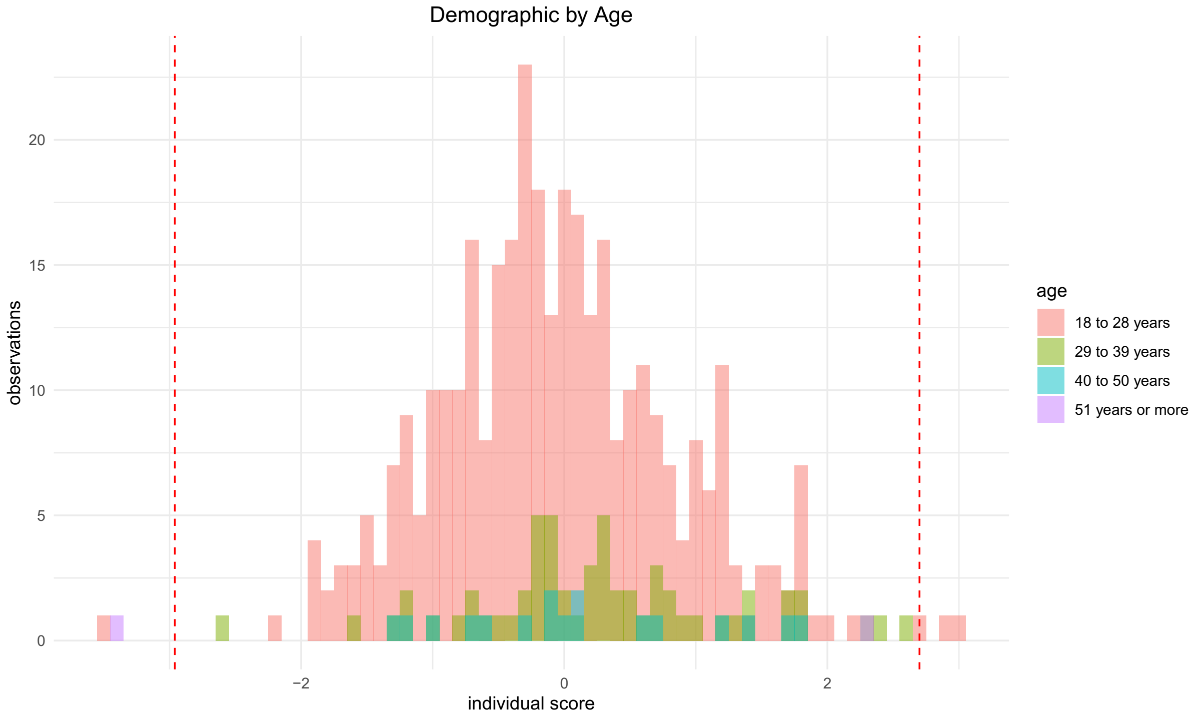
As shown in Figures 32 to 37, the normal curves generated for each group were significantly close, indicating that our model did not present different behaviors in the observed groups and the cumulative probability distribution. Furthermore, the sigmoid function suggests that the model does not significantly differentiate discrimination parameters (a – the sigmoid slope) and difficulty (b – the sigmoid midpoint).

In the demographic facet of participants' age (Figure 32), the normal distribution and the sigmoid function do not present a score distortion in 3 of the four 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 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.

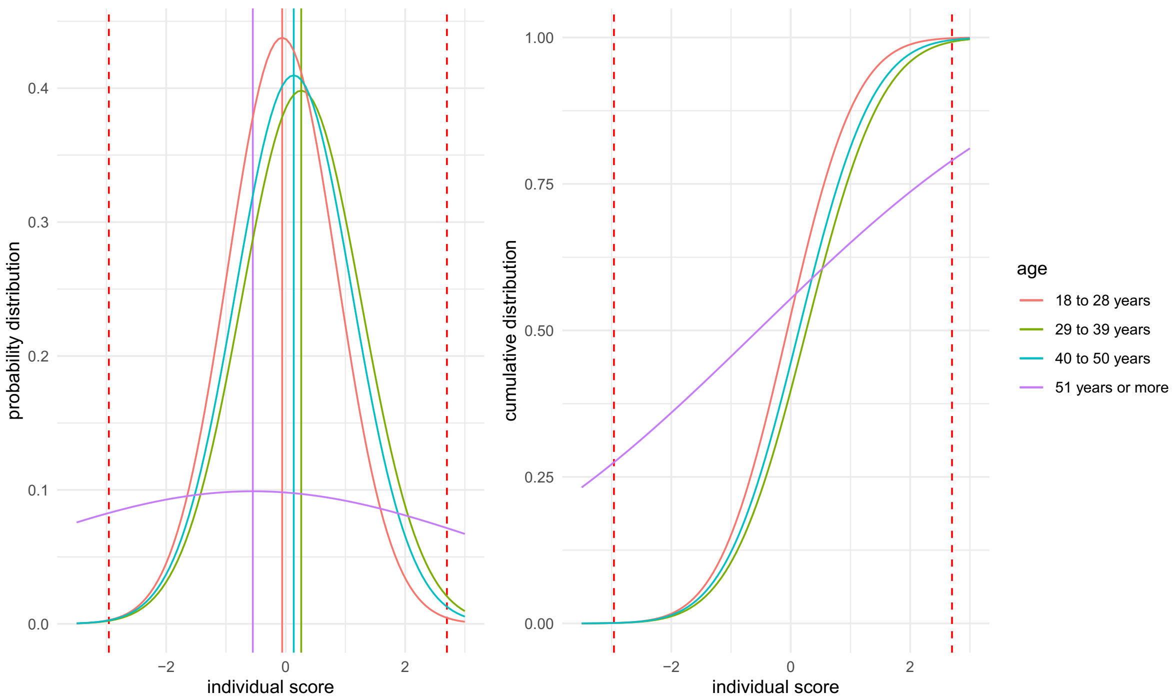
Grouping the participants by gender (Figure 33), we demonstrated that the model does not present additional difficulties or differentiate participants depending on their gender options. Furthermore, despite the sample mainly comprising males, females, and "other genders" participants obtained similar results in both mean scores and normal and sigmoid functions.

Comparing the scores grouped by the preferred videoconferencing environment (Figure 34), 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. Despite the larger number of observations collected for Google Meet, Microsoft Teams, and Moodle groups, due to the IRT scale invariance property[1], we argue that the IRT items' calibration for other environments presents the same representativeness. Therefore, possible distortions in the scale due to the proportionality of different demographic groups were considered.

[1] Due "item-free" and "person-free" estimation invariance (AYALA, 2013a), the skills/abilities and item parameters being estimated based on the responses of a group of individuals, once the skill measurement scale is established, the item parameters do not change. Consequently, IRT values are invariant to different groups of respondents as long as individuals in these groups have their skills measured on the same scale (ANDRADE; TAVARES; VALLE, 2000).

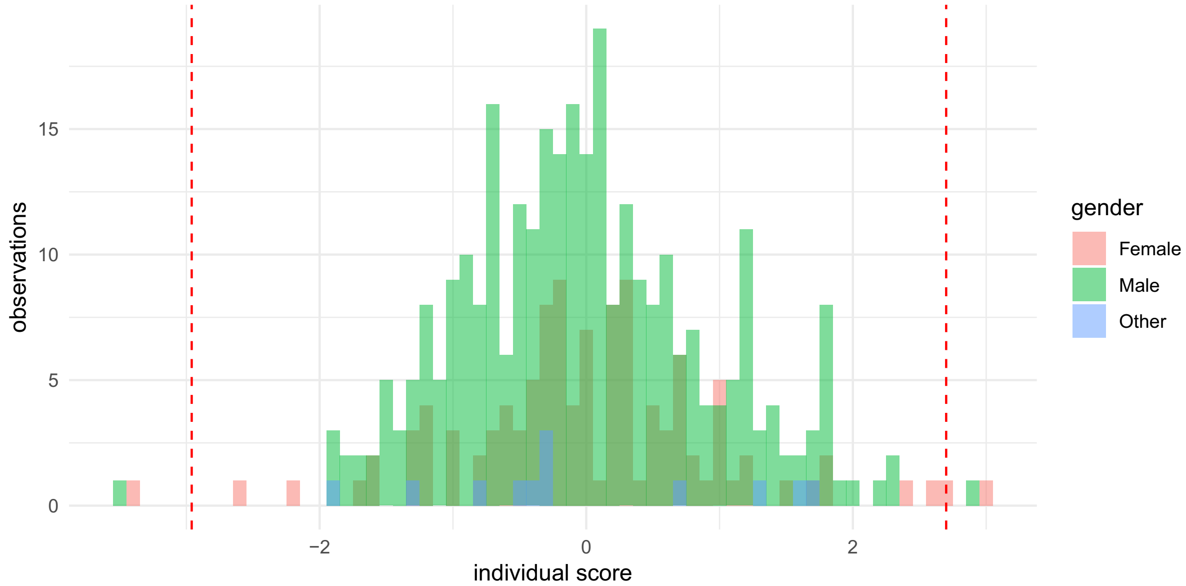


(a) Score

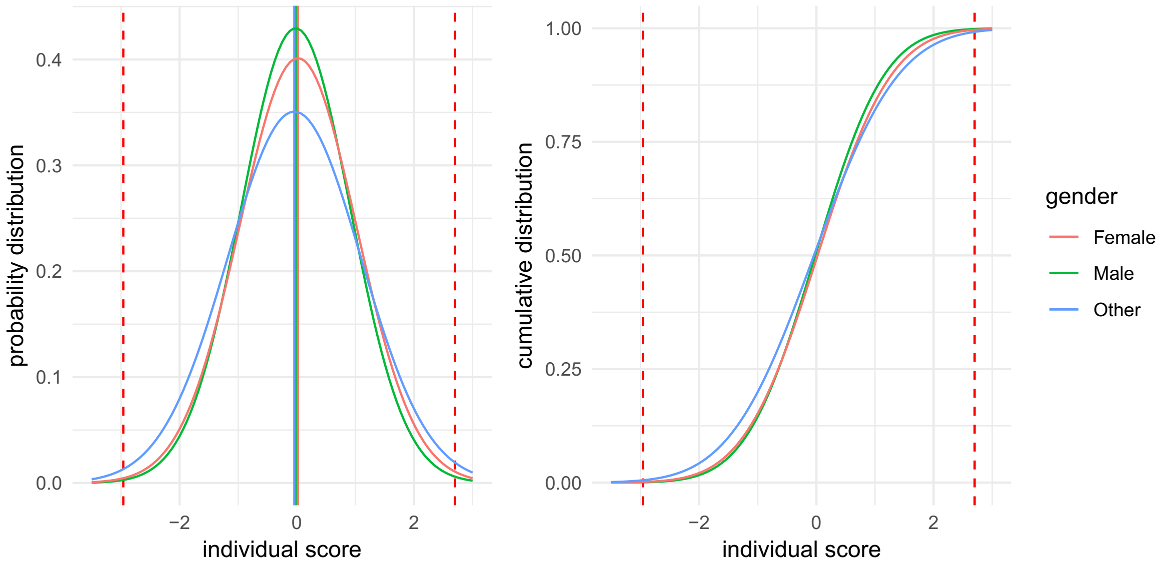


(b) Normal curves

Figure 32 – Demographic distribution of individual score by age (videoconferencing scenario)

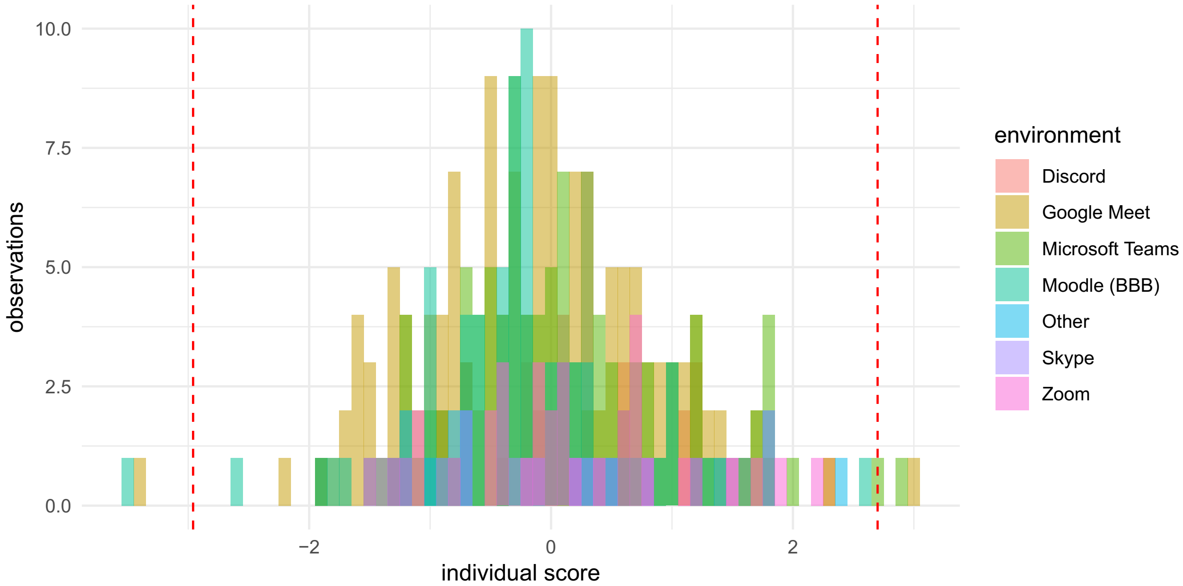


(a) Score

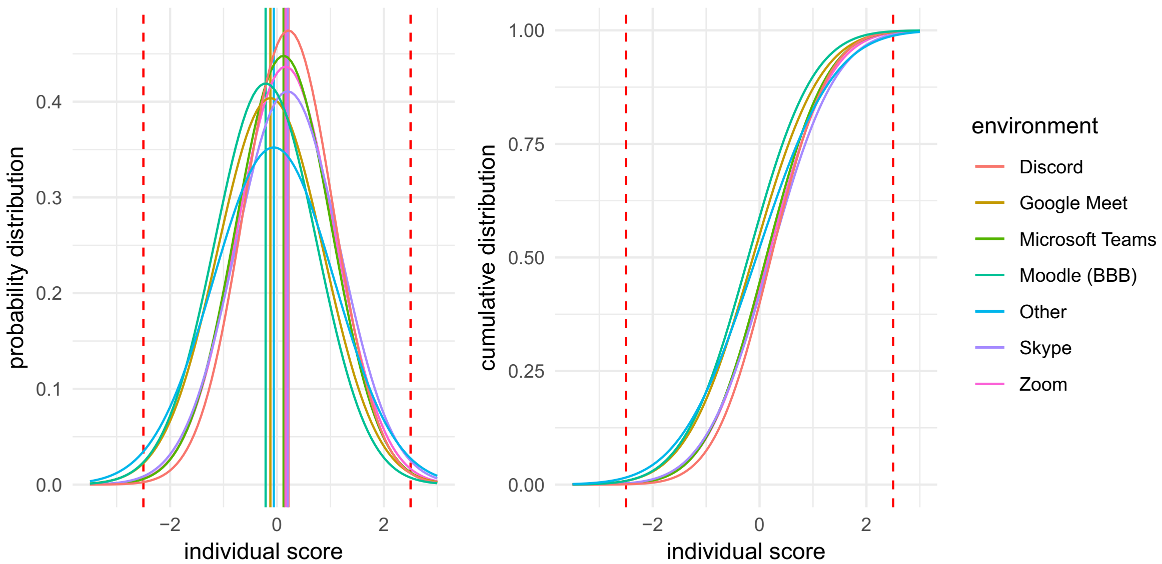


(b) Normal curves

Figure 33 – Demographic distribution of individual score by gender (videoconferencing scenario)

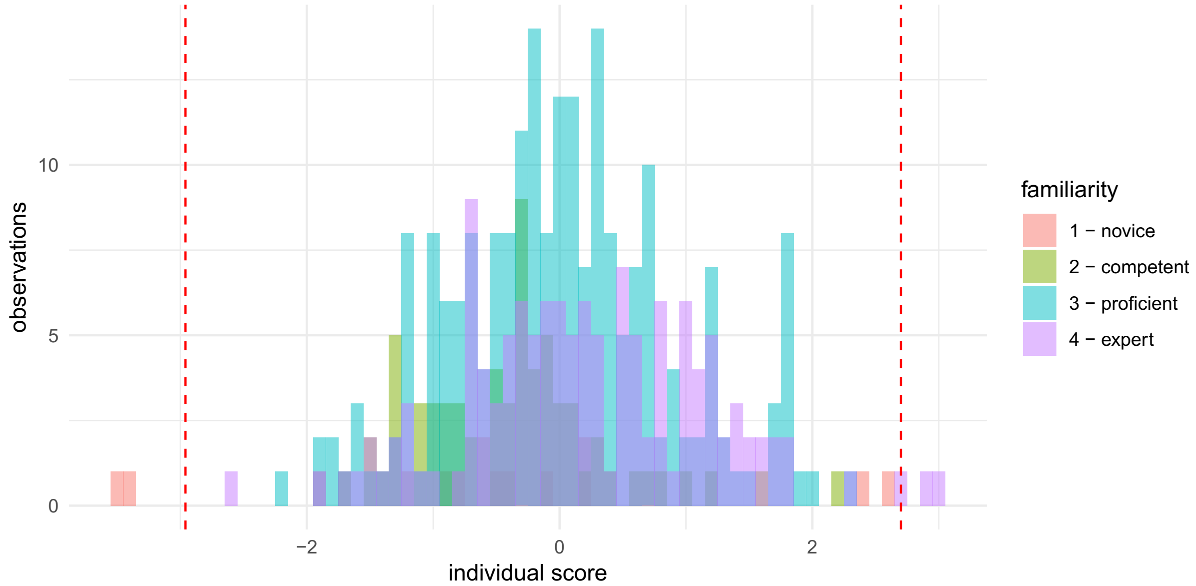


(a) Score

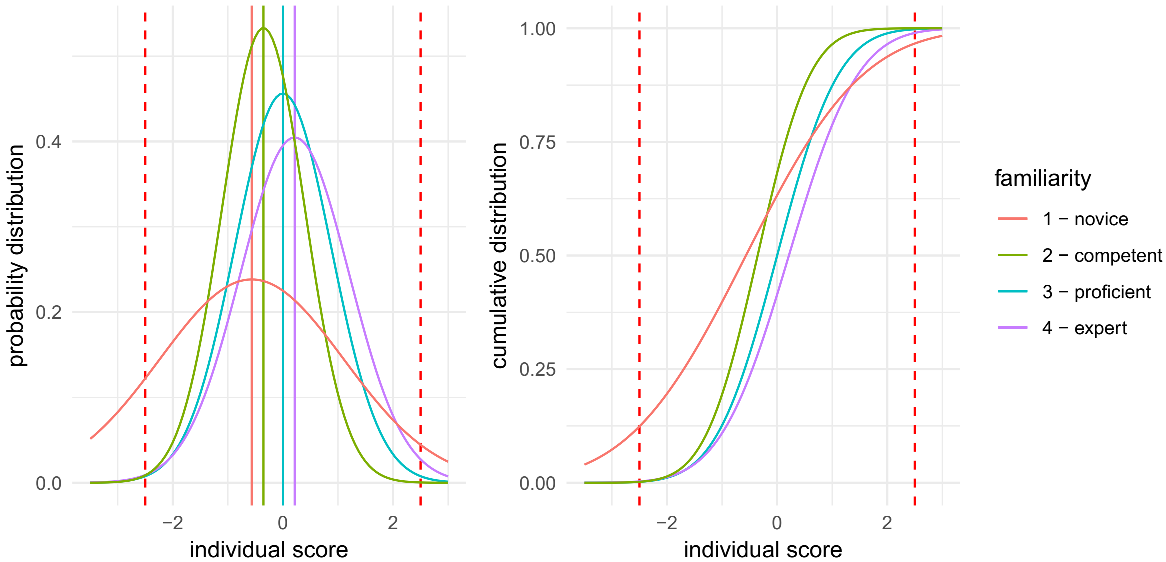


(b) Normal curves

Figure 34 – Demographic distribution of individual score by environment (videoconferencing scenario)

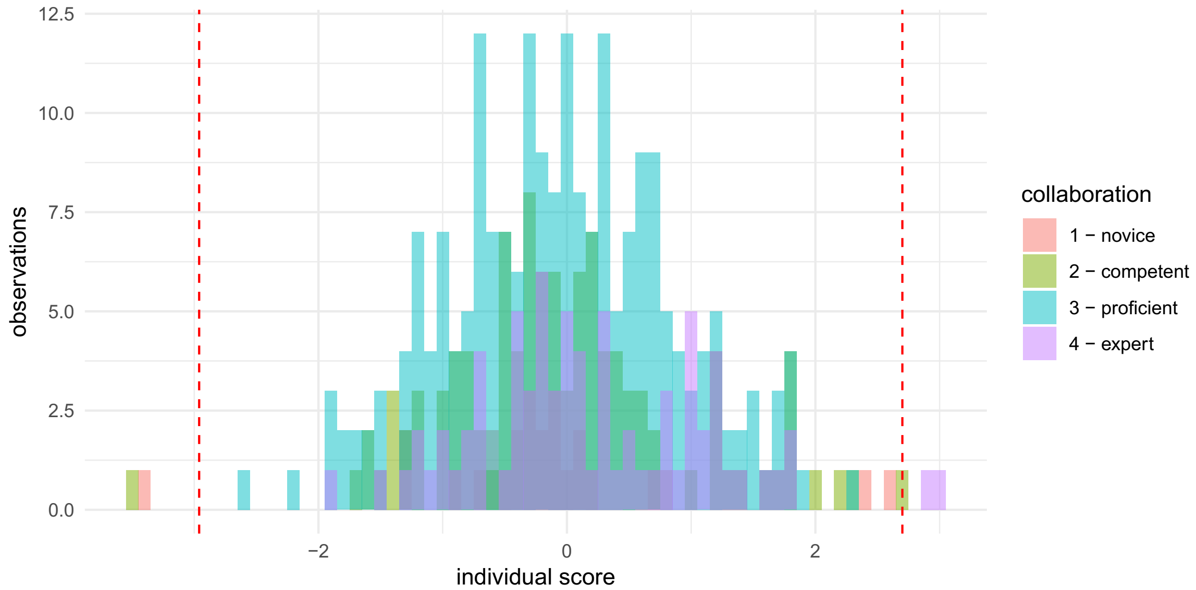


(a) Score

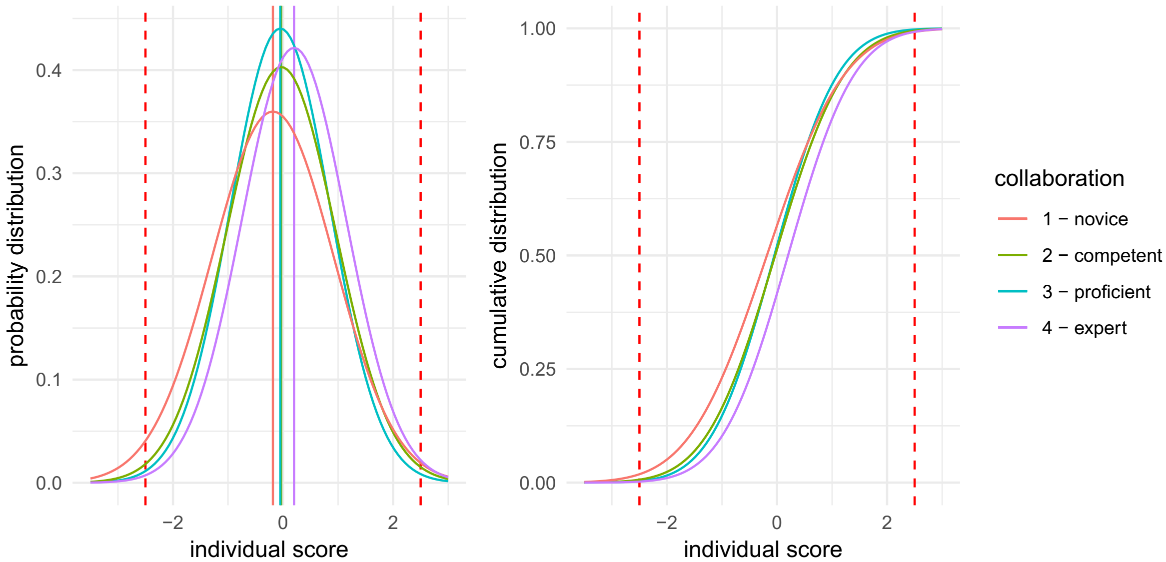


(b) Normal curvesy

Figure 35 – Demographic distribution of individual score by familiarity (environment) (videoconferencing scenario)

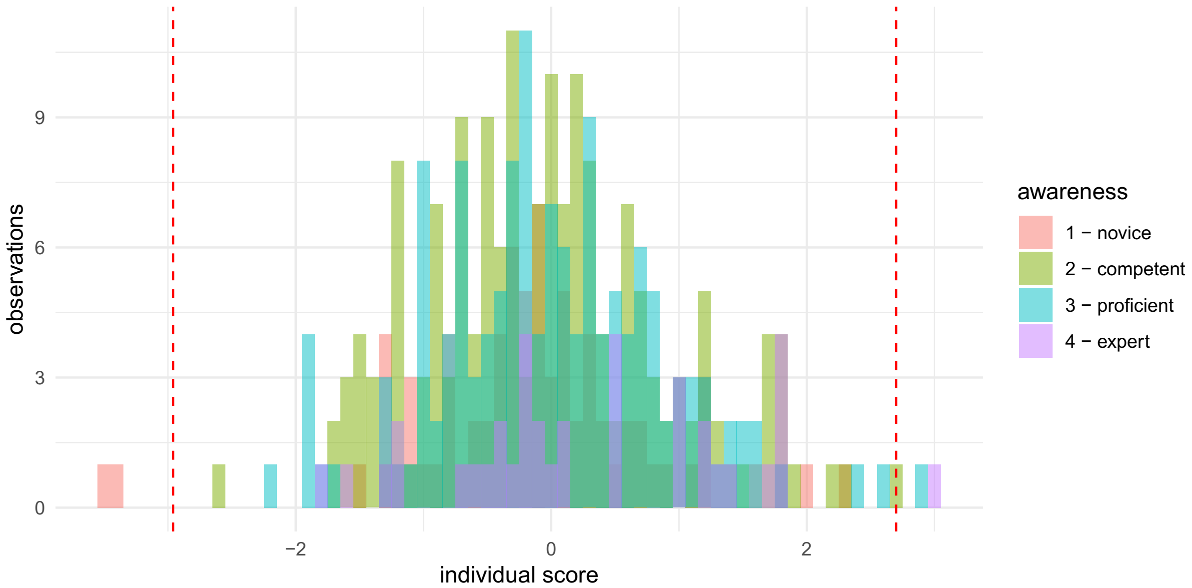


(a) Score

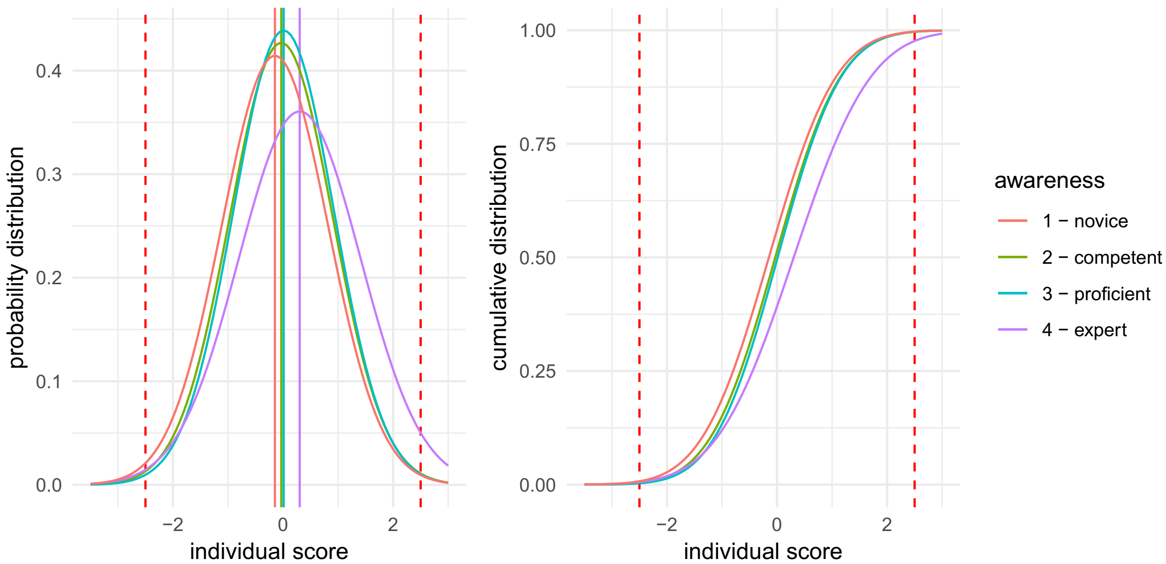


(b) Normal curves

Figure 36 – Demographic distribution of individual score by familiarity (collaboration) (videoconferencing scenario)



(a) Score



(b) Normal curves

Figure 37 – Demographic distribution of individual score by familiarity (awareness) (videoconferencing scenario)

By analyzing the participants' individual skill histograms (Figures 35a, 36a, and 37a), whether familiarity with the preferred videoconferencing environment, collaboration, or awareness concepts, both normal distribution and probability cumulative distribution (Figures 35b, 36b, and 37b) 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 Figures 32 to 37, 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 our taxonomy (described in Section 4), 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 38 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 (THISSEN; WAINER, 2001); the red dotted line represents the standard error $SE(\theta)$. The intersection point represents the limits at which the model is more representative.

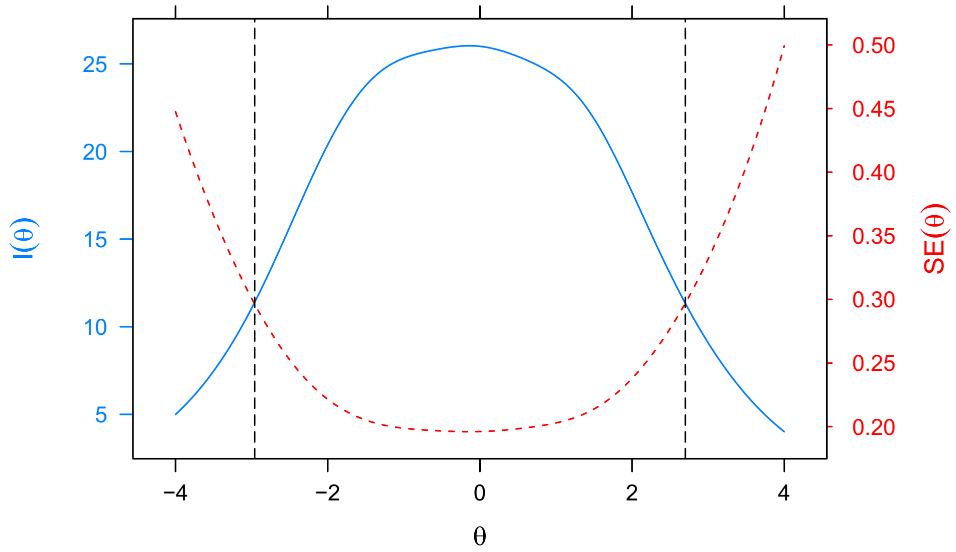


Figure 38 – Test information and SE

This graph represents the region of the ability scale θ_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 $[-2.96, +2.70]$.

9.2.3 Assessing the awareness support scale

Applying the awareness measurement formulas 15 to 17 defined in Section 5.7.1, we calculated the probability scales $P_{i,k}(\theta_j)$ for the assessment element through our IRT awareness assessment model.

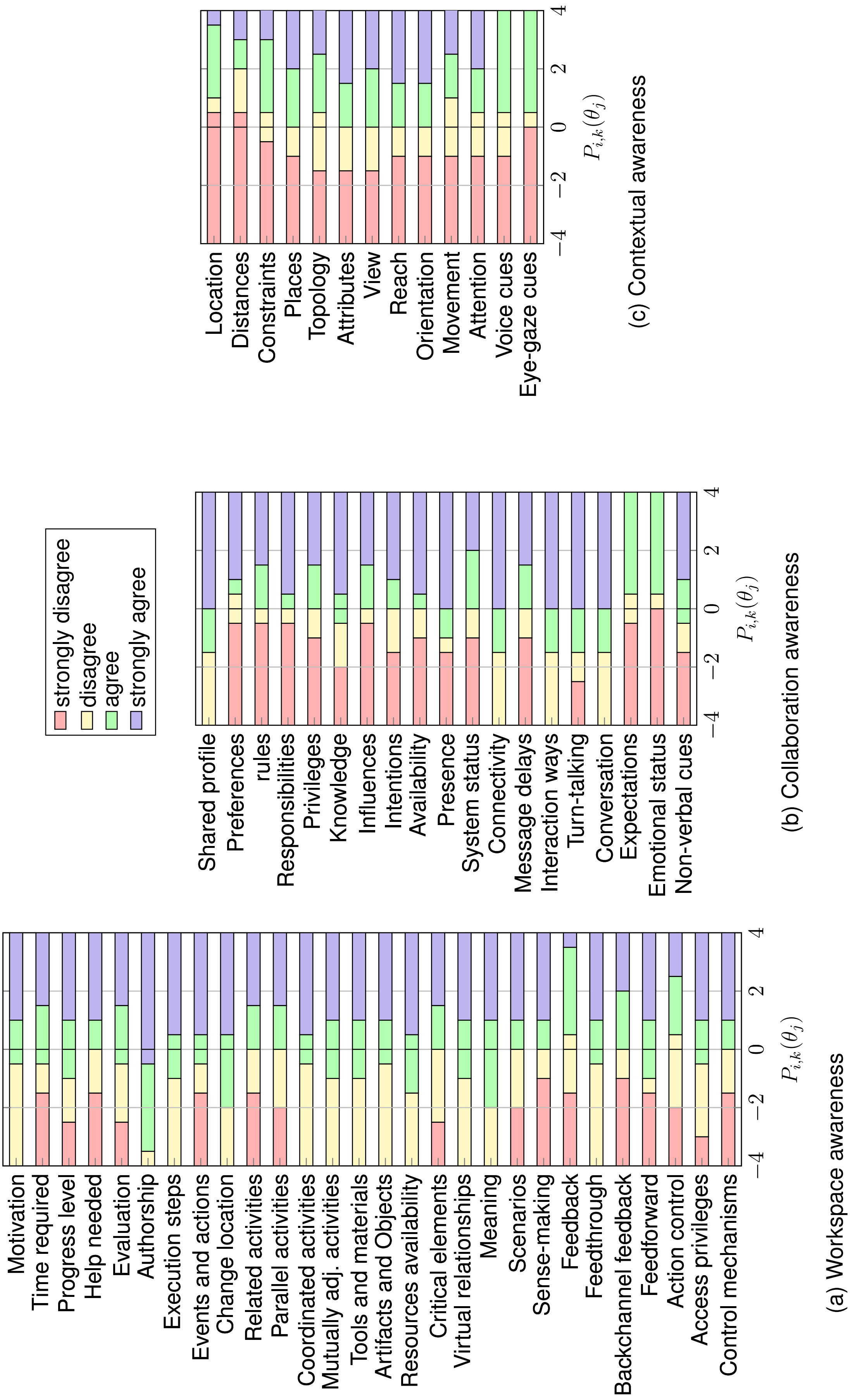
In our awareness support scale, we assume a coverage interval $[-4.0, +4.0]$, although our model is representative at the interval $[-2.96, +2.70]$, to cover the outlier scores of individuals with lower or higher abilities ($< 1\%$). Figure 39 presents the probability scales generated for each assessment 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 θ can assume any real value between $-\infty$ and $+\infty$. Therefore, it is necessary to establish an origin and a unit of measurement to define the scale (ANDRADE; TAVARES; VALLE, 2000).

To calibrate our model and construct the graphs shown in Figures 38 to 39, we considered the scale with a mean μ equal to 0, and a standard deviation σ 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 (ANDRADE; TAVARES; VALLE, 2000).

Despite the frequent use of this $(0, 1)$ scale, there are no practical differences if these or any other values μ and σ are established; just the order relationships between their points matter (PASQUALI, 2020). 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.



(a) Workspace awareness

(b) Collaboration awareness

(c) Contextual awareness

Figure 39 – Ability level scales (videoconferencing scenario)

To overcome a possible scale misinterpretation, we adopted the principle of invariance of the IRT scales (BAKER, 2001; BAKER; KIM, 2017) and applied a linear transformation (Equation 18) to establish a more appropriate and easier reference for people to interpret their awareness score through a positive scale θ^*.

$$a(\theta - b) = \frac{a}{\sigma}[(\sigma.\theta + \mu) - (\sigma.b + \mu)] = a^*(\theta^* - b^*) \tag{18}$$

with,

$$a^* = \frac{a}{\sigma}; \qquad b^* = \sigma.b + \mu;$$
$$\theta^* = \sigma.\theta + \mu; \qquad P(\theta^*) = P(\theta)$$

where,

$*$ indicates the value at the new scale;
μ is the mean θ value at the first scale;
σ is the standard deviation at the first scale;
θ^* is the adjusted score (the awareness points);

Over the participant's score θ and IRT params a and b, we applied a linear transformation converting the resultant scores to a new scale $(100, 10)$. In this perspective, we positioned the calibrated items over the awareness scale and established 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 θ and θ^* 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 31, 32, 33 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 for both to identify how awareness works and how the collaboration takes place in these environments. Essentially, this model provides reflections toward collaborative improvements by gradually prioritizing supported awareness elements over a participant's perspective.

Table 31 – Workspace awareness scales (videoconferencing scenario)

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 some 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 32 – Collaboration awareness scales (videoconferencing scenario)

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 in 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 33 – Contextual awareness scales (videoconferencing scenario)

Level	Quality description
Low $(\theta < -1)$ $(\theta^* < 90)$	The collaborative environment hardly considers the contextual perspective nor the group members' mobility. Navigation or spatiality are rarely allowed and the 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.

9.3 PHASE 3: REFLECTION

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

In the first awareness scale perspective, as shown in Figure 39, 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), we represent the expected ability intervals in which participants present a certain probability P_i of selecting each response category presented. In other words, starting from the participants' ability scale θ, we represent the probable intervals $P_i(\theta)$ that participants are most likely to correctly identify/understand the awareness mechanism in the evaluated interface.

As we can see in 39c, 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 31, 32, and 33, we categorized the results concerning the skill levels of the expected participants. In our assessment model, the ability scale θ encompasses within the interval $[-4.0, +4.0]$ and the adjusted ability scale θ^* encompasses within the interval $[+60, +140]$ (awareness points). Then, we established three participant ability intervals, describing the expected competencies concerning the awareness mechanisms participants in each ability score interval understand.

9.3.1 Scale interpretation

In our 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 acquire the more important awareness mechanisms easily. In short, the higher the participant's degree of ability, the better the score obtained in the assessment of environmental awareness support.

From the workspace awareness perspective, hardly awareness support mechanisms are received for participants with a minor individual ability ($\theta < -2$)[2]. In this videoconferencing scenario, we can observe more positive feedback on the assessment compared to the previous case study (Moodle/e-learning supporting environment), and a more significant result can be viewed both in the workspace and in the collaboration awareness perspective.

Time required, help needed, events and actions, related and parallel activities, sense-making, feedback, backchannel feedback, feedforward, control mechanisms, and less markedly, progress level, evaluation, critical elements, scenarios, action control, and access privileges elements demanded a greater ability to be recognized in the environment (generally for individuals with score θ below -2). Some awareness mechanisms, like motivation, authorship, execution steps, change location, coordinated and mutually adjusted activities, tools and materials, artifacts and objects, resources availability, virtual relationships, meaning, and feedthrough, did not present great difficulty in identifying them (strongly disagree) – although participants with a lower level of ability reported certain disagreement with these resources.

On the other hand, for ability scales close to $\theta = 0$, most participants agreed, or even strongly agreed, with the availability of workspace awareness mechanisms. This indicates that, from this ability level, videoconferencing environments generally present an adequate configuration of resources to support collaborative work, and the feeling of the shared workspace necessary for interaction can be captured. Furthermore, by ability $\theta = 1$, participants have a great tendency to strongly agree with these resources, indicating that from this ability level onward, participants can take advantage of these mechanisms satisfactorily.

From the collaboration awareness perspective, shared profile, connectivity, interaction ways, and conversation mechanisms did not present great difficulty in its appropriation; even at lower scores (like $\theta < -1$), participants rarely strongly disagree with these resources. All other collaboration awareness mechanisms demanded a greater

[2] This represents novice participants below two standard deviations from the observed mean ($\theta = -2$).

ability to be recognized in the environment, and participants with less ability ($\theta < -1$) generally strongly disagreed with these resources.

Participants begin satisfactorily absorbing some collaboration aspects from the ability $\theta = -1$ onward, achieving a better understanding of the awareness resources at skill levels close to $\theta = 0$. Shared profile, knowledge, presence, connectivity, interaction ways, turn-talking, conversation, and non-verbal cues are examples of easier awareness mechanisms for participants at abilities between $-1 \leq \theta \leq 0$.

Similarly to workspace awareness elements, from ability $\theta = 0$ onward, most participants agreed or strongly agreed with the availability of collaboration awareness mechanisms. This indicates that, from this ability level, videoconferencing environments generally present an adequate configuration of resources to support this awareness perspective. Furthermore, by ability $\theta = 1$, participants tend to strongly agree with these resources, indicating that from this ability level onward, participants can take advantage of these mechanisms satisfactorily. On the other hand, concerning expectations and emotional status mechanisms, participants were unlikely to strongly agree with these elements, which suggests difficulties or noise in their understanding; as a recommendation, a thorough analysis of collaboration resources should be considered.

From the contextual awareness perspective, we can identify a marked difficulty in understanding the awareness mechanisms evaluated; however, the results were slightly better than those identified in the Moodle scenario. For participants with lower ability scores ($\theta \leq -1$), contextual mechanisms are unlikely to be captured, and for scores between $-1 \leq \theta \leq 0$, participants generally tend to disagree with these resources.

Places, attributes, view, reach, and orientation mechanisms begin to be perceived when the ability is above $\theta = 0$, and the others when θ is closer to 1. Finally, participants tend to strongly agree with these resources only at high scores ($\theta \geq +2$). Voice cues and eye-gaze cues are two mechanisms that are difficult to strongly identify, even in scores at the upper end of the scale (θ closer to $+4$). In conclusion, from a contextual point of view, the evaluated environment presents a great lack of support that must be considered.

In summary, the results corroborate our understanding awareness assumption that it is intrinsically linked to participants' skills in identifying, understanding, or projecting their actions: *i)* Awareness requires consideration of the individual's abilities in recognizing and projecting next actions in the environment; and *ii)* There is a link between mapping the awareness mechanism's availability and the participant's understanding (knowledge, mental state, previous experiences, etc).

9.4 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 (α)[3].

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 (θ)'s scale scores. We use the IRT technical properties of discrimination (a) and difficulty (b), combined with α coefficient, to assess the Awareness Assessment Model Instrument reliability and internal consistency.

9.4.1 Reliability results

Both alpha and IRT params strongly demonstrate the validation of our proposed model. First, the adequate representation of the awareness scale θ (interval $[-2.96, +2.70]$ as presented in Figure 38) is good evidence of the instrument's reliability. In addition, the internal reliability through Cronbach's alpha coefficient (DEVELLIS, 2016) demonstrated an excellent internal consistency for all assessment items $(\alpha > 0.91)$. Second, we calculate the instrument's reliability function $r_{xx}(\theta)$ (THISSEN; WAINER, 2001; CHALMERS, 2012), across participants' latent trail (see Figure 40). Our 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.

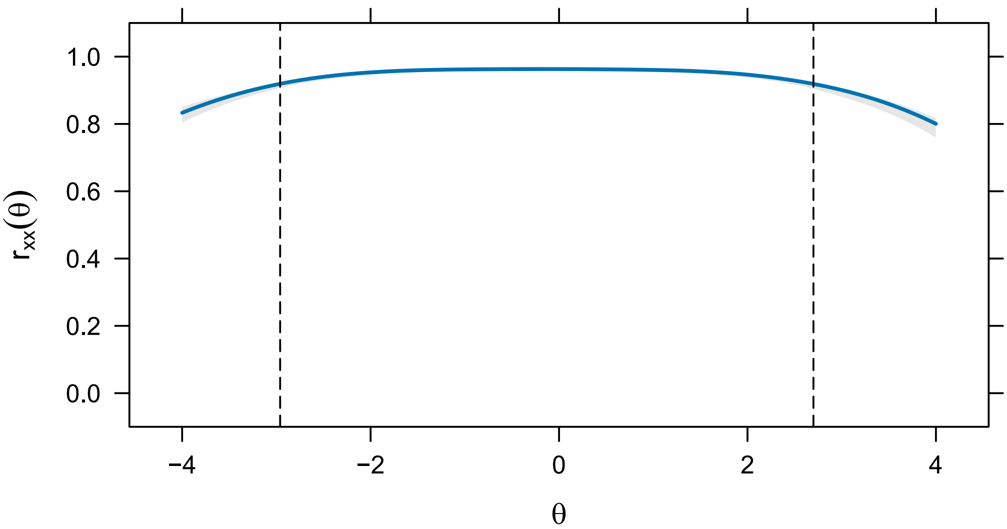


Figure 40 – Test reliability (videoconferencing scenario)

[3] 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 (DEVELLIS, 2016).

9.5 MODEL DIMENSIONALITY

An essential factor that corroborates the IRT model's validation is the latent trait's dimensionality, which, in our case, refers to the number of factors necessary to explain the variability of the data and constitute a hypothesis to be verified (SINGH, 2004). 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 (ANDRADE; TAVARES; VALLE, 2000). This factor is supposed to be measured by the assessment instrument. Schmitt (1996) 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 (SPENCER, 2004).

We used the Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) to test dimensionality (SINGH, 2004). 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 explain the correlations pattern (OSBORNE, 2014). 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 (THOMPSON, 2004; BROWN, 2015).

Due to the sample size, we used the same data set for the EFA and CFA. In this configuration, Izquierdo, Olea, and Abad (2014) 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 (HAIR et al., 2009). In the Latent Root, the factors or components retained in the analysis with real data must have an eigenvalue higher than ones obtained randomly (LEDESMA; VALERO-MORA, 2019); 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 (HAIR et al., 2009).

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 (PASQUALI; PRIMI, 2003). 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 (PASQUALI; PRIMI, 2003). 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 (HAIR et al., 2009).

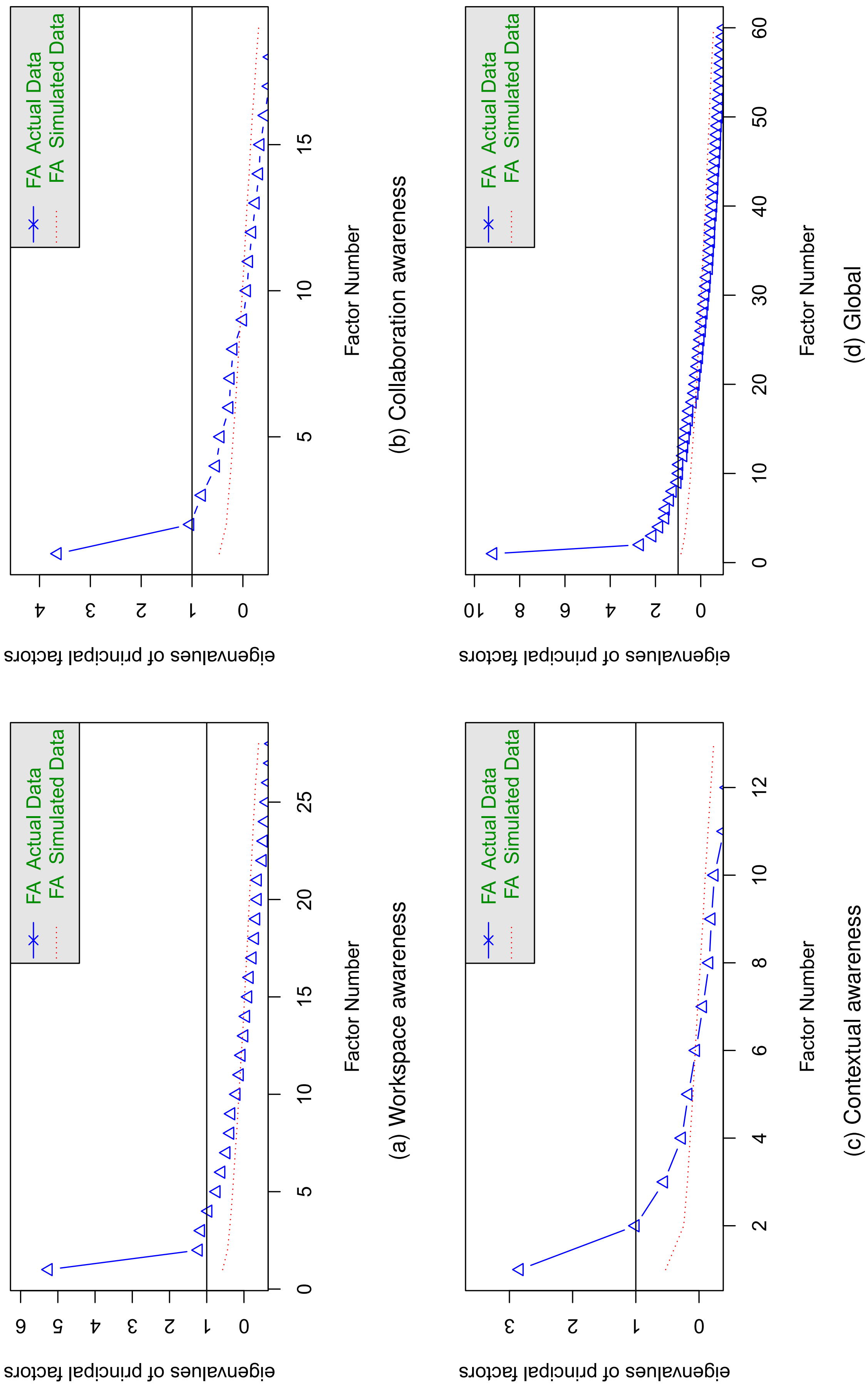
9.5.1 EFA results

Factor analysis is based on the simulation of random data to determine the number of factors (IZQUIERDO; OLEA; ABAD, 2014). The factors/components retained with real data must have an eigenvalue higher than those obtained randomly.

The latent root criterion suggests a strong principal component and three other prominent components. As we can see in Figures 41a, 41b, and 41c, by applying the Kaiser criterion (eigenvalue ≥ 1), we identify three principal components for the workspace, two 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.

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 collaboration awareness perspective, Factor 1 explains 23.07% ($pc = 3,5431$); and in the contextual awareness perspective, Factor 1 explains 27.25% ($pc = 4,3831$) of the total variance. Tables 34 to 36 present the total variance explained by each principal component (pc). The literature suggests that the 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 (RECKASE, 1979).



(a) Workspace awareness

(b) Collaboration awareness

(c) Contextual awareness

(d) Global

Figure 41 – Model dimensionality (videoconferencing scenario)

Table 34 – Total variance explained: workspace awareness (videoconferencing scenario)

pc	Total	Variance	Cumulative		pc	Total	Variance	Cumulative
1	6.0161	21.49%	21.49%		15	0.6760	2.41%	83.97%
2	2.0596	7.36%	28.84%		16	0.6523	2.33%	86.30%
3	1.9762	7.06%	35.90%		17	0.5970	2.13%	88.43%
4	1.8215	6.51%	42.40%		18	0.5411	1.93%	90.36%
5	1.5607	5.57%	47.98%		19	0.4908	1.75%	92.12%
6	1.4184	5.07%	53.04%		20	0.4089	1.46%	93.58%
7	1.2820	4.58%	57.62%	...	21	0.3773	1.35%	94.92%
8	1.2012	4.29%	61.91%		22	0.3471	1.24%	96.16%
9	1.1388	4.07%	65.98%		23	0.2929	1.05%	97.21%
10	0.9739	3.48%	69.46%		24	0.2591	0.93%	98.13%
11	0.9439	3.37%	72.83%		25	0.2116	0.76%	98.89%
12	0.8603	3.07%	75.90%		26	0.1494	0.53%	99.42%
13	0.8281	2.96%	78.86%		27	0.0911	0.33%	99.75%
14	0.7546	2.70%	81.55%		28	0.0704	0.25%	100.00%

Table 35 – Total variance explained: collaboration awareness (videoconferencing scenario)

pc	Total	Variance	Cumulative		pc	Total	Variance	Cumulative
1	4.3832	23.07%	23.07%		11	0.6745	3.55%	83.85%
2	1.8733	9.86%	32.93%		12	0.6466	3.40%	87.25%
3	1.6382	8.62%	41.55%		13	0.5552	2.92%	90.18%
4	1.3832	7.28%	48.83%		14	0.4546	2.39%	92.57%
5	1.2385	6.52%	55.35%	...	15	0.4405	2.32%	94.89%
6	1.1308	5.95%	61.30%		16	0.3484	1.83%	96.72%
7	1.0483	5.52%	66.82%		17	0.2956	1.56%	98.28%
8	0.9629	5.07%	71.89%		18	0.2348	1.24%	99.51%
9	0.8714	4.59%	76.47%		19	0.0928	0.49%	100.00%
10	0.7272	3.83%	80.30%					

9.5.2 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 (HAIR et al., 2009). Based on the EFA results, we re-wrote the awareness perspectives based on three factors for the workspace, two for the collaboration, and two for the contextual perspectives.

Table 36 – Total variance explained: contextual awareness (videoconferencing scenario)

pc	Total	Variance	Cumulative		pc	Total	Variance	Cumulative
1	3.5431	27.25%	27.25%		8	0.6469	4.98%	84.08%
2	1.8468	14.21%	41.46%		9	0.5751	4.42%	88.51%
3	1.3777	10.60%	52.06%		10	0.4930	3.79%	92.30%
4	1.0900	8.38%	60.44%	...	11	0.4412	3.39%	95.69%
5	0.9168	7.05%	67.50%		12	0.3242	2.49%	98.19%
6	0.8595	6.61%	74.11%		13	0.2356	1.81%	100.00%
7	0.6500	5.00%	79.11%					

Figures 42, 43, and 44 contains the graphical representation of the model, correlating the factors and their related awareness elements (assessment items) and the factor loadings. A factor loading above 0.30 indicates a moderate correlation between the item and the factor (PASQUALI; PRIMI, 2003; TAVAKOL; WETZEL, 2020). Extracting the evidence from the factor analysis of the model, we identified the main latent dimensions (factors) for the workspace and two for collaboration and contextual awareness perspectives, as described in Tables 37, 38, and 39.

We could visualize a significant equivalence by comparing the model generated by the EFA analysis of the assessment items of our taxonomy (awareness elements and design categories). Furthermore, as shown in Figures 42, 43, and 44, most factor loadings exceed 0.3. In only two cases, the factor loading found was below this cutoff value. Therefore, we opted to maintain the corresponding evaluation items (Q31 – access privileges and Q49 – message delays) since the calibration of the item parameters was satisfactory and, from the CFA perspective, new scenarios can investigate their reliability in the awareness support assessment.

The confirmatory factor analysis results largely maintained the structure defined in our taxonomy, and the factor loadings of the CFA items demonstrate the instrument's construct validity. We also calculated Composite Reliability (CR) to evaluate the construct validity of the proposed model (HAIR et al., 2009). The measurements obtained were then evaluated following the recommendations of Fornell and Larcker (1981). All factors in the model must present a CR value above 0.7 to demonstrate the instrument's construct validity (FORNELL; LARCKER, 1981; HAIR et al., 2009).

Evidence of construct validity indicates that the items measured in the sample represent the real measurements in the population (HAIR et al., 2009). From the workspace awareness perspective, factors F1, F2, and F3 presented CR values of 0.712, 0.692 ($\simeq 0.7$), and 0.706. Regarding collaboration awareness, factors F1 and F2 presented CR values of 0.825 and 0.761, respectively. From the contextual awareness perspective, values of 0.708 and 0.715 were found for factors F1 and F2.

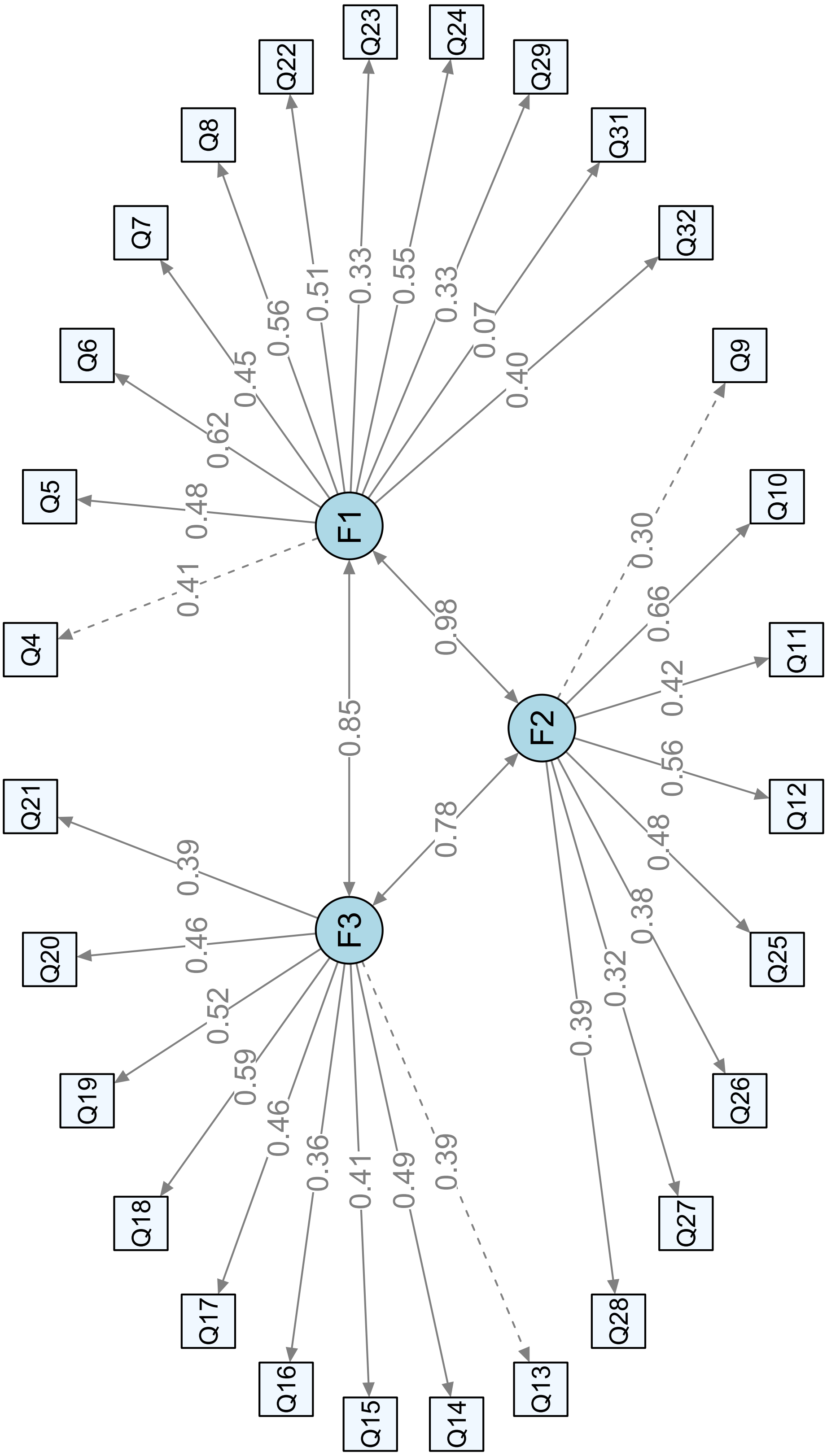


Figure 42 – Workspace awareness (videoconferencing scenario)

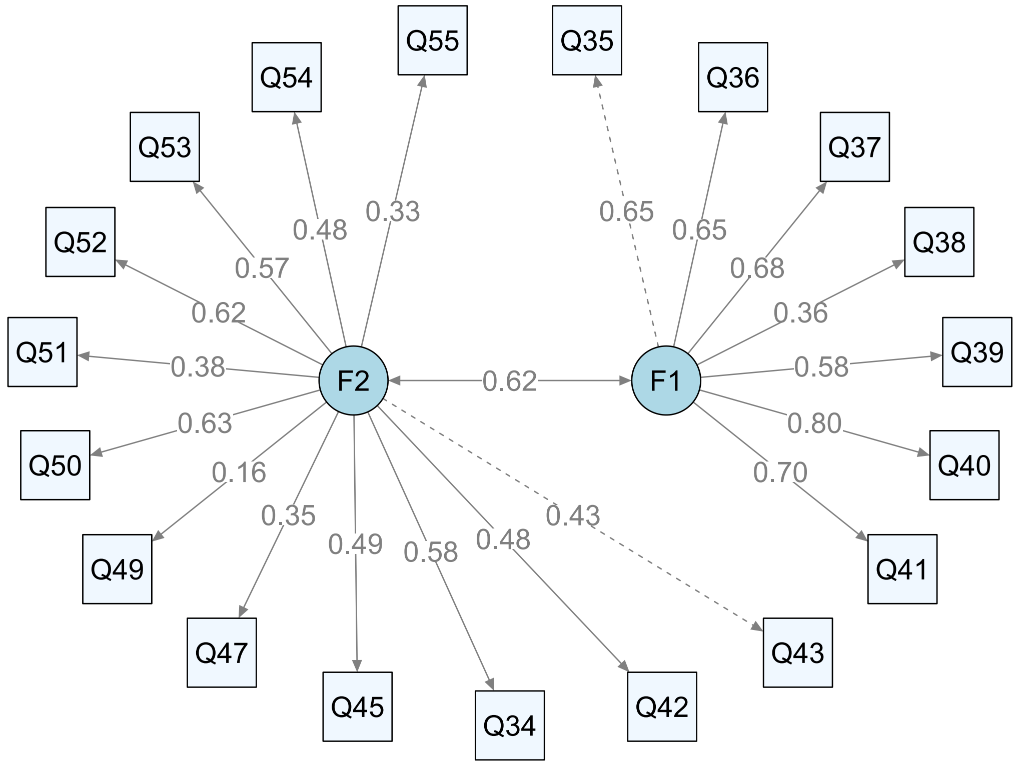


Figure 43 – Collaboration awareness (videoconferencing scenario)

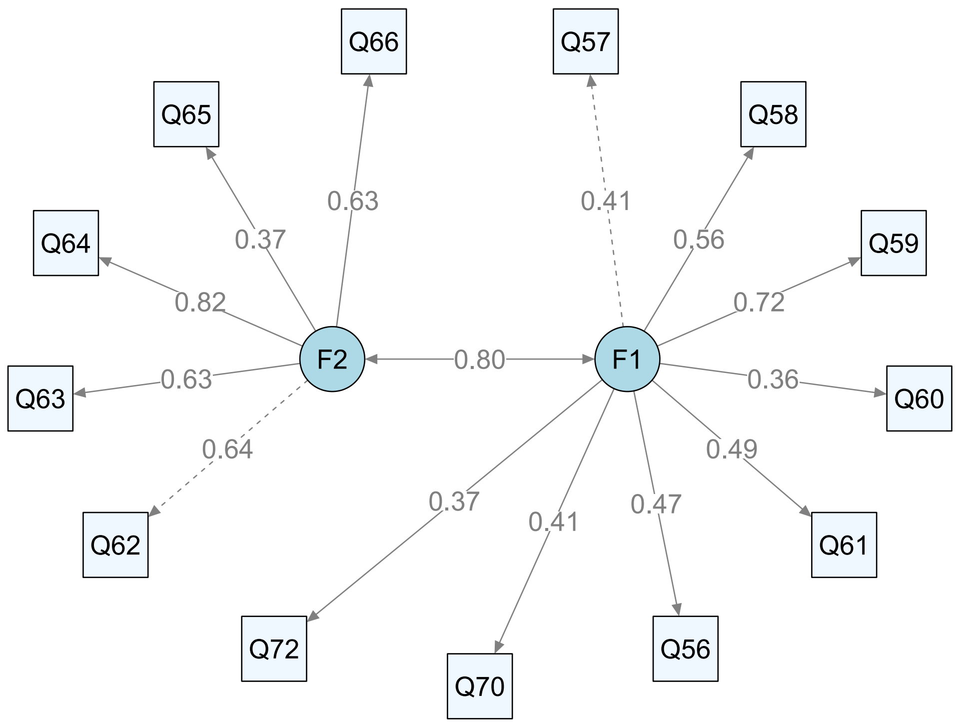


Figure 44 – Contextual awareness (videoconferencing scenario)

Table 37 – Factor analysis of the workspace awareness perspective (videoconferencing scenario)

Factor	Factor description
F1 (Factor 1)	It 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)	It 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 feedforward feedback).
F3 (Factor 3)	It 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 38 – Factor analysis of the collaboration awareness perspective (videoconferencing scenario)

Factor	Factor description
F1 (Factor 1)	It 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)	It 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-talking, conversation, expectations, emotional status, and non-verbal cues).

Table 39 – Factor analysis of the contextual awareness perspective (videoconferencing scenario)

Factor	Factor description
F1 (Factor 1)	It represents 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)	It represents 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).

9.6 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 smaller overall awareness score, respectively, θ equals to -0.21, -0.12, and 0.12; the adjusted ability scores θ^* 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, θ equals to 0.16, 0.20, and 0.21; the adjusted ability scores θ^* 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 θ, 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 believed that non-calibration occurred due to many positive (agree or strongly agree) or negative (disagree or strongly disagree) responses. In the first, our 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 θ, we calculated the probability $P_{i,k}(\theta_j)$ considering the Samejima's gradual scale (SAMEJIMA, 1969). 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.

In all 60 calibrated assessment items, θ, 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[4].

We used the exploratory factor analysis (EFA) and the confirmatory factor analysis (CFA) to test the model dimensionality. The EFA results indicated a strong tendency towards the one-dimensional model (latent root criterion) (HAIR et al., 2009), legitimating the correlation between the assessment items and the observed latent trail. The CFA results demonstrate the instrument construct validity: all factors presented adequate composite reliability (CR ≥ 0.7) (FORNELL; LARCKER, 1981; HAIR et al., 2009), and most all factor loadings of the assessment items are above the limit (≥ 0.3) (TAVAKOL; WETZEL, 2020; PASQUALI; PRIMI, 2003).

9.6.1 Observed 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 θ of a universe of 60 calibrated assessment items, each with four response categories – Samejima's gradual scale (SAMEJIMA, 1969). 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)

[4] 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 (DEVELLIS, 2016).

(LINACRE, 1999; WIRTH; EDWARDS, 2007). This grouping may eventually affect the item calibration.

The IRT model could not calibrate some items from the 75 assessment items initially proposed in our taxonomy, as discussed in section 9.2.1. Thus, validating these items was also impossible in this work; alternatively, we suggest a global evaluation scenario to investigate their suitability for accessing awareness support in collaborative environments – this will be covered in Chapter 11.

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 our repository (MANTAU; BENITTI, 2023).

9.6.2 Related publications

The results presented in this chapter were published in:

- MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. The Awareness Assessment Model: Measuring Awareness and Collaboration Support Over Participant's Perspective. In: Universal Access in the Information Society (UAIS). DOI: 10.1007/s10209-024-01110-5.
- MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. The Awareness Assessment Model repository. Version 1.0. Zenodo, Aug. 2023. DOI: 10.5281/zenodo.8298950.

10 CASE STUDY 3: THE ASSESSMENT PROCESS VALIDATION

Notably, representative results are expected from a good evaluation model or process. Conversely, hard-to-use models or processes obscure the reliability between the obtained result and the observed object. We believe that the success of a good evaluation is related to the rigor of the model (e.g., artifacts, questionnaires, analysis spreadsheets, and synthesis available) and the evaluation process conducted; thus, both must be accessed for their reliability and usefulness.

According to Nickerson, Varshney, and Muntermann (2013) and Szopinski, Schoormann, and Kundisch (2019), usefulness is related to the purposeful, unambiguous determination, and applicability aspects. In this scenario, purposeful is the relevance of the assessment process (significance and objectivity of its elements/activities); unambiguous determination is the process's correctness and understandability (ability to represent its elements and activities concisely and unambiguously); and applicability refers to the process's authenticity, generality, usability, and concreteness attributes.

For usefulness assessment, we expose the model's artifacts to HCI and collaborative system examiners' appreciation to verify the suitability of the process, its activities, and related artifacts to evaluate the awareness support in collaborative environments.

10.1 SCENARIO

This scenario consists of two parts. First, examiners adopted the model artifacts and process to evaluate the awareness support in general-purpose collaborative office tools (e.g., most common text editing tools, spreadsheets, and document managers). Second, the model's conceptual view artifacts and evaluation process usefulness were assessed through researcher observations and questionnaire.

We selected mostly novice examiners in HCI and collaborative system application evaluations to identify potential difficulties in replicating the proposed model. Thus, we sought to evaluate whether novices see the practical use of the model and its artifacts in the same way as the previous model's expert panel validation indicated (Chapter 7). Examiners were accompanied during the artifact preparation and assessment process activities.

Before starting the case study, participants were invited to a briefing in which the evaluation model and their artifacts were presented. At this stage, all materials necessary for the study, including a detailed overview of the evaluation process, the awareness taxonomy, the inventory of assessment questions, a template of collection instruments (questionnaires, printed and online), and a report model of the assessment results were exposed to all examiners groups.

This scenario included 25 examiners (19 males and six females) divided into seven assessment groups. The sample comprises undergraduate computing students with a basic notion of HCI, software quality, and software process concepts. It took three meetings (2 hours each) to complete the evaluation activities. Finally, participants were invited to respond to the usefulness assessment questionnaire – similar to that used in the expert panel (Chapter 7).

The group configuration was as follows: Groups 1 and 2 carried out the evaluation of the Google Sheets environment, collecting 50 and 36 observations respectively; Groups 2 to 6 evaluated the Google Docs environment, totaling 42, 49, 45, and 12 observations respectively. Group 7 evaluated the Trello environment, collecting 15 observations.

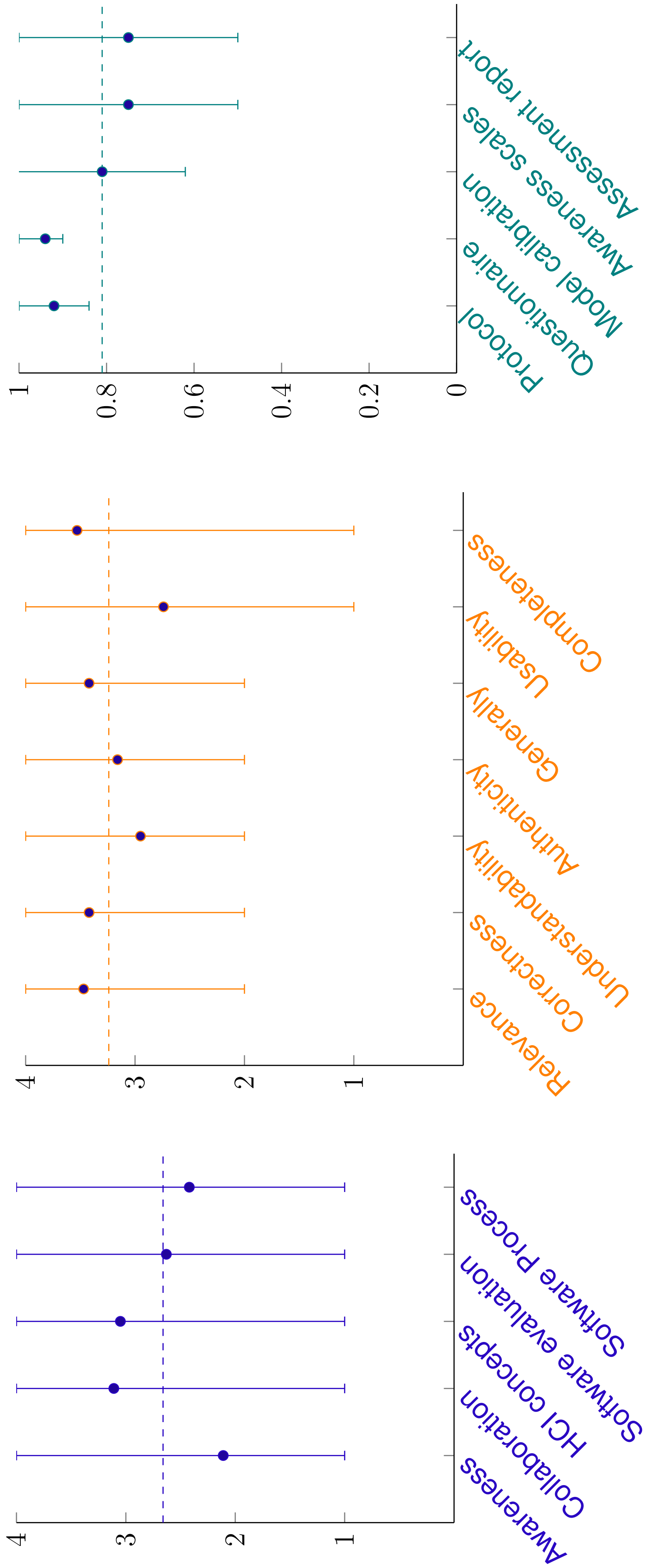
In total, 249 observations(157 males, 90 females, and two did not respond) were collected. The observations collected in each scenario will be incorporated into our global analysis, described in Chapter 11. All artifacts, including the Awareness Assessment Model templates, the artifacts generated from each team, and the demographic and utility questionnaire collected, are available in the model repository (MANTAU; BENITTI, 2023).

10.1.1 Results

We presented the demographic and usefulness questionnaire to the participants, and 19 responses were obtained (76%). Six examiners did not answer the questionnaire. Although a small sample was obtained, the assessment model received a good rating from the examiners' viewpoint. Figure 45 summarizes the results in three basic facets: demographic, usefulness, and assessment artifacts.

On the demographic facet (Figure 45a), the examiners have a varied knowledge of awareness, collaboration, HCI, software process, and software evaluation. On a gradual scale, from 1 (novice) to 4 (expert), the average reported expertise in these related concepts was 2.67, indicating a reasonable familiarity with this context. Awareness, software process, and software evaluation concepts were the least familiar aspects to the examiners (respectively, average 2.11, 2.63, and 2.42). Due to the examiners' sample variability, experience with key concepts ranged across the spectrum of the gradual scale.

Although a small sample, the model's insight into the different skill levels was found. Furthermore, we observed that most of the difficulties in applying the model are related to the participant's skills in awareness concepts and statistical processes. Some reported not having in-depth knowledge about the evaluated tools; thus, planning the assessment (assessment protocol artifact) took a while, compromising time for other activities, such as data collection and analysis.



(a) Demographic

(b) Usefulness

(c) Artifacts

Figure 45 – Assessment process validation results

On the usefulness facet (Figure 45b), the examiners indicated a good evaluation of both the model (artifacts, questionnaires, analysis spreadsheets, and synthesis available) and the evaluation process conducted. On a gradual scale, from 1 (strongly disagree) to 4 (strongly agree), the average usefulness reported was 3.24. The relevance, correctness, generality, and completeness aspects were significantly well evaluated (close to 3.5). Comprehensibility and authenticity obtained an average rating close to 3 (agree). Usability presented the lowest value in the usefulness assessment, pointing to 2.74.

The usefulness assessment results show that the purposeful perspective, captured by relevance, can be satisfactorily met, even for examiners unfamiliar with the related key concepts. Likewise, from the unambiguous determination point of view, represented by correctness and comprehensibility elements, the evaluated model also presented good results; the model presents considerable clarity and consistency in its artifacts and assessment activities, which can be applied without major difficulties by examiners, even those unfamiliar with the approach.

The model's applicability perspective, which considers authenticity, generality, usability, and completeness, presented an interesting result. All items were evaluated well, demonstrating that the model gives a true domain account and indicates if the reference model is usable in different evaluation scenarios. Regarding usability, examiners generally demonstrated that they could operate, implement, and apply the reference model easily, although they were supervised throughout the process. Regarding completeness, examiners indicated that the representation contains all statements about the domain that are correct and relevant. Only one of the respondents strongly disagreed with the model's usability and completeness.

We also evaluated the artifacts constructed by each group of examiners to verify their elaboration and identify potential difficulties. Materials like model artifacts, raw data, spreadsheets, results, and other necessary auxiliary files were analyzed. We then assign a score from 0 (none) to 1 (excellent) for the quality of the generated material. Figure 45c) summarizes the results. We identified five main artifacts generated: assessment protocol, questionnaire (data collection instrument), IRT model calibration, awareness scales definition, and assessment report. Overall, the groups elaborated artifacts with considerable quality, obtaining an overall average score of 0.81.

The calibration of the model, construction of awareness scales, and assessment report, on the other hand, presented varied results. Although all presented an acceptable average general quality (above 0.75), some groups presented difficulties during preparation. A first point of emphasis is that due to the limited sample size that each group obtained (between 15 and 75 observations), the calibration and subsequent steps were hampered. Furthermore, the knowledge of the group of examiners related to the key concepts necessary to conduct the assessment was varied (see Figure 45a).

10.2 DISCUSSION

The groups present minor difficulties regarding: *a)* unfamiliarity with the IRT and HCI assessment; *b)* model complexity, *c)* statistical analysis (novices); *d)* short time available to assimilate the resources of the target tool, select the categories and awareness mechanisms, and construct the data collection instrument (questionnaire) based on the chosen mechanisms.

In the cases (*a*) and (*c*), it is imperative to highlight that for a proper HCI assessment, the examiner's prior knowledge of the tool and the processes adopted, and the assessment itself is crucial. This applies to IRT concepts and basic statistical knowledge. Therefore, we relaxed the analysis of this aspect since our interest in this evaluation involved the assessment model replication in other scenarios, contexts, and examiners – even those with little knowledge of the analyzed facets (like context, target tool, or awareness, collaboration and HCI concepts) (see Figure 45a).

The awareness assessment model was designed to encapsulate part of the natural complexity of IRT and statistical analysis, presenting some analysis and assessment scale templates alongside the model. However, it may be difficult to apply this model to examiners as their first contact with one statistically based HCI assessment model (*b*). At this point, the analysis of the model's complexity has been hampered, and broader scenarios can be considered.

Similarly, appropriating awareness concepts and assessment elements was necessary in the second case (*d)*. Examiners with more favorable knowledge about awareness and collaboration concepts obtained the best results at this stage due to the short time available. As examiners explored the context, target tool, and assessment process, they easily identified and selected the design categories and awareness mechanisms and constructed the data collection instruments (*c*); both artifacts presented an excellent overall average quality above 0.9.

This scenario demonstrated that the model can be replicated fully or partially in other scenarios and contexts by selecting the dimensions, categories, and awareness mechanisms relevant to the scenario and adapting the awareness support scale. Due to the small sample size and examiners' knowledge of key concepts, new assessment scenarios may be required to verify the required, or even recommended, knowledge for examiners to replicate the model properly.

Part IV

Global Awareness Scale

11 THE GLOBAL AWARENESS SCALE

In the IRT model, one of the most important steps is estimating items' parameters and participants' abilities; this process is known as calibration (ANDRADE; TAVARES; VALLE, 2000). The calibration can be done by knowing the items' parameters, by knowing the participant's abilities, or by estimating both simultaneously (most common situation). In the first strategy, only the abilities can be estimated; in the second, only the items' parameters are estimated.

In this chapter, we present a global awareness scale based on data obtained through scenarios of case studies 1 to 3 (Chapters 8 to 10). We assume the estimation of both participants' abilities and items' parameters as a calibration strategy and the IRT multi-group estimation method (CHALMERS, 2012).

We estimate the multiple-group calibration by performing a maximum likelihood analysis for polytomous data (Samejima's gradual scale (SAMEJIMA, 1969)) using the Metropolis-Hastings Robbins-Monro (MHRM) algorithm approach (CAI, 2010).

To calibrate the model and generate the global awareness scale, we considered each of the scenarios as a distinct evaluation group: case study 1 (group 1), case study 2 (group 2), and case study 3 (groups 3 to 9). In this scenario, two or more groups take two or more tests, only partially different (with some common items). In this configuration, the common items between different tests allow all parameters to be on the same scale at the end of the estimation processes. As a set of items connects the different populations, it is possible to make comparisons and build a global scale (ANDRADE; TAVARES; VALLE, 2000).

To estimate the participants' skills in different scenarios, it is necessary to equalize the items' parameters, that is, to make comparisons or place the parameters from different tests or the skills of respondents from different populations on the same metric or common scale (ANDRADE; TAVARES; VALLE, 2000). IRT equating is the statistical instrument used to compare different test scores from different forms when IRT models assemble tests, allowing the scores from both tests to be used interchangeably.

According to Holland, Dorans, and Petersen (2006), numerous data collection designs have been used for score equating. In this global scale scenario, we combined the equivalent groups (EG) strategy and the Equating by common item procedure (ANDRADE; TAVARES; VALLE, 2000). In the EG design, two or more equivalent samples are taken from a common population P; each P_i is tested (HOLLAND; DORANS; PETERSEN, 2006); then EG is performed by taking random samples from $P = (P_1, P_2, \ldots, P_n)$. In equalization via common items, the guarantee that the involved populations will have their parameters on a single scale will be provided by common items between the populations, which will link them (ANDRADE; TAVARES; VALLE, 2000).

11.1 MODEL CALIBRATION

We compiled all 820 observations obtained through all evaluation scenarios (described in Chapters 8 to 10) into a .csv file to generate the Global Awareness Scale. We executed the IRT script available in the assessment model repository (MANTAU; BENITTI, 2023). Similarly to the case studies presented in Chapters 8 and 9, we interpreted the IRT values of discrimination (a) and difficulty (b), disregarding items with $a < 0.65$ or $a > 4.0$. We also analyzed the observed frequencies of each response category, and for all questionnaire items, the observed frequency was ≥ 10.

Tables 40 to 42 present the coefficients of discrimination (a) and difficulty (b), the observed frequencies, and Cronbach's alpha coefficient (α). Appendix F, Figures 86 to 88 summarizes the items' information functions for each awareness dimensions. The coefficients $b1$, $b2$, and $b3$ are related to the responses category disagree, agree, and strongly agree.

In the workspace awareness perspective (see Table 40), all assessment items present values compatible with the range defined for the parameters a and b. In the collaboration awareness perspective (see Table 41), item Q33 - identity did not converge to satisfactory parameters. We conjecture that it may indicate that this assessment item is strongly linked to user-specific factors because participants generally indicated ease in identifying it and chose positive responses (agree or strongly agree statements). From the contextual awareness perspective (see Table 42), Q68 - user mobility was removed from the results, indicating that this resource was absent or had not been used by the participants to collaborate.

The model calibration converged to satisfactory discrimination and difficulty coefficients for all assessment items. Additionally, the internal consistency values obtained through Chronbac's alpha coefficient indicate excellent model reliability[1].

By analyzing the accumulated frequency of each response category, we observed that the sample obtained presents good variability between all skill scales of individuals who disagree or agree with each of the mechanisms evaluated. Therefore, it was not necessary to adopt the grouping of response categories. As a result, on our global awareness scale, we could represent the participant's general position in judging the item's existence and the degree of agreement in the response (e.g., agree or strongly agree, and disagree or strongly disagree).

[1] 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 (DEVELLIS, 2016).

Table 40 – Workspace awareness coefficients and observed frequencies (global scale)

Our Awareness Taxonomy		Item	Items' coefficients (a and b)				Observ. frequencies gradual scale (1 to 4)				Alpha (α)
			a	$b1$	$b2$	$b3$	**1**	**2**	**3**	**4**	
Involves activities	Goal	Q1	0,7084	-4,2342	-2,1701	0,9022	18	41	128	102	0,9342
	Subject	Q2	0,8607	-4,1476	-1,8306	0,7459	15	64	183	155	0,9338
	Content	Q3	0,8928	-3,6713	-2,1031	0,3704	16	34	129	136	0,9337
	Motivation	Q4	0,9781	-3,5894	-1,1867	1,4707	10	57	119	45	0,9345
	Time required	Q5	1,0054	-2,7095	-0,6795	1,4609	21	69	112	61	0,9342
	Progress level	Q6	0,9503	-2,9252	-1,0287	1,1474	30	84	161	110	0,9336
	Help needed	Q7	1,4907	-2,0836	-0,4088	0,9671	25	95	107	68	0,9334
	Evaluation	Q8	1,1440	-2,8004	-0,8941	1,1575	20	76	141	78	0,9334
Consider workflow	Authorship	Q9	0,9485	-3,5501	-1,7508	0,0089	18	52	120	200	0,9333
	Execution steps	Q10	1,5761	-2,4145	-0,9574	0,8212	13	49	134	82	0,9335
	Events and actions	Q11	1,7062	-2,2389	-0,9186	0,6698	18	70	169	121	0,9333
	Change location	Q12	0,8187	-3,6344	-1,6514	0,8126	18	46	108	89	0,9340
	Related activities	Q13	1,4680	-2,0870	-0,5919	1,3397	23	72	132	40	0,9338
	Parallel activities	Q14	1,2501	-0,8007	0,4533	2,1224	85	82	83	34	0,9332
	Coordinated activities	Q15	1,5894	-2,0418	-0,5035	0,7828	20	69	85	57	0,9335
	Adjusted activities	Q16	1,4291	-2,7817	-0,9695	1,2306	10	46	121	51	0,9338
Consider environment	Tools and materials	Q17	1,0852	-3,2775	-1,1457	1,1703	14	69	160	94	0,9334
	Artifacts and objects	Q18	1,3332	-2,5600	-0,7936	1,1293	14	59	125	65	0,9339

continues on the next page

Table 40: Workspace awareness coefficients and observed frequencies (Global Awareness Scale) (continuation)

	Resources availability	Q19	0,9690	-3,3744	-1,4898	0,8572	17	53	135	98	0,9337
	Critical elements	Q20	1,0117	-2,8106	-0,3588	1,7652	23	96	108	44	0,9339
	Virtual relationships	Q21	1,6082	-2,7171	-0,7726	0,9168	10	55	103	53	0,9332
Provide understanding	Meaning	Q22	1,0530	-3,2250	-1,5797	0,9764	15	43	155	91	0,9339
	Scenarios	Q23	1,2185	-2,2148	-0,2995	1,3348	25	88	103	62	0,9334
	Sense-making	Q24	1,7324	-1,4864	0,0726	1,4060	39	105	89	46	0,9332
Allow interaction	Feedback	Q25	0,9259	-3,8483	-1,4460	0,7293	12	66	139	125	0,9335
	Feedthrough	Q26	1,2676	-2,7814	-0,7137	1,2402	14	76	147	78	0,9335
	Backchannel feedback	Q27	1,3752	-1,4328	0,1353	2,0191	46	97	99	26	0,9337
	Feedforward	Q28	1,2843	-2,2558	-1,0280	0,7826	39	79	196	145	0,9332
Consider relationship	Action control	Q29	0,9575	-1,8047	0,4066	2,6963	50	106	84	23	0,9344
	Access control	Q30	0,8826	-3,5731	-1,7428	0,8540	14	36	111	84	0,9345
	Access privileges	Q31	0,9606	-2,7211	-0,7616	1,1361	29	80	116	87	0,9340
	Control mechanisms	Q32	1,6612	-1,9208	-0,3280	1,0686	27	117	159	104	0,9329

Table 41 – Collaboration awareness coefficients and observed frequencies (global scale)

Our Awareness Taxonomy		**Item**	**Items' coefficients (a and b)**				**Observ. frequencies** gradual scale (1 to 4)				**Alpha** (α)
			a	$b1$	$b2$	$b3$	**1**	**2**	**3**	**4**	
Allow identity	Identity	Q33	–	–	–	–	32	69	163	211	–
	Shared profile	Q34	0,9910	-3,2687	-1,6987	0,6395	22	52	164	131	0,9334
	Preferences	Q35	1,9149	-0,9143	0,3950	1,6915	56	106	74	27	0,9328
Consider capabilities	Rules	Q36	1,2317	-1,4718	-0,0456	1,9159	61	95	120	39	0,9338
	Responsibilities	Q37	2,0729	-1,2139	-0,0599	1,2706	40	88	105	45	0,9332
	Privileges	Q38	1,1823	-1,6277	0,1929	1,8394	71	144	112	50	0,9337
	Knowledge	Q39	1,2538	-2,2042	-0,4614	1,2935	30	84	103	57	0,9339
	Influences	Q40	1,7560	-1,3702	-0,0355	1,6464	37	87	108	31	0,9334
	Intentions	Q41	1,8456	-1,6645	-0,1183	1,5100	24	91	113	35	0,9333
Provide status	Availability	Q42	1,4319	-1,5175	-0,0901	1,1097	43	84	79	69	0,9335
	Presence	Q43	0,8350	-2,7266	-1,1946	0,6774	38	57	110	136	0,9341
	Activity level	Q44	0,8295	-2,3544	0,1522	2,1383	45	119	100	58	0,9339
	Status	Q45	1,1507	-1,9166	-0,1083	1,7003	45	110	115	48	0,9336
Provide communication	Mode	Q46	0,7094	-3,6759	-1,1335	1,3347	19	55	87	69	0,9347
	Connectivity	Q47	1,0325	-2,8266	-1,4676	0,3241	16	33	79	91	0,9346
	Message delivery	Q48	0,8988	-2,6075	-1,0069	0,9676	42	76	137	112	0,9343
	Message delays	Q49	0,7540	-1,9406	-0,0713	1,9989	61	78	82	50	0,9343
	Interaction ways	Q50	1,0045	-2,9781	-1,6841	0,6990	23	40	140	101	0,9339
	Turn-talking	Q51	0,9616	-2,6422	-0,9979	0,4132	27	56	78	123	0,9341
	Conversation	Q52	0,7324	-3,4939	-1,5727	1,0753	31	65	156	126	0,9340
Consider social	Expectations	Q53	1,4192	-1,3913	0,5060	2,0164	63	152	84	28	0,9336
	Emotional status	Q54	1,2026	-0,5440	0,8646	2,6519	114	91	58	14	0,9336
	Non-verbal cues	Q55	0,7859	-1,8615	-0,1072	1,7230	65	86	94	76	0,9338

Table 42 – Contextual awareness coefficients and observed frequencies (global scale)

Our Awareness Taxonomy		Item	Items' coefficients (a and b)				Observ. frequencies gradual scale (1 to 4)				Alpha (α)
			a	$b1$	$b2$	$b3$	**1**	**2**	**3**	**4**	
Consider spatiality	Location	Q56	0,8863	-0,3491	1,4523	3,4363	122	89	50	18	0,9342
	Distance	Q57	1,0245	-0,4139	1,9935	2,9067	92	97	16	14	0,9347
	Restrictions	Q58	1,5217	-0,9372	0,6871	1,9658	57	99	45	18	0,9344
	Places	Q59	1,2212	-1,9462	-0,1902	2,0605	32	90	124	32	0,9332
	Topology	Q60	1,1659	-2,1367	-0,3514	2,0259	34	86	121	30	0,9339
	Attributes	Q61	1,6421	-1,6001	-0,0298	1,7319	27	84	86	22	0,9335
	View	Q62	1,0314	-1,9180	-0,1768	1,8828	49	106	142	65	0,9338
	Reach	Q63	0,9849	-1,4360	0,9359	2,6606	69	128	58	21	0,9338
	Orientation	Q64	1,3789	-1,3743	0,5080	2,3479	57	123	75	16	0,9334
	Movement	Q65	1,0834	-1,9285	-0,1214	1,6002	56	127	129	78	0,9336
	Range of attention	Q66	1,6901	-1,4523	0,1112	2,0077	45	111	98	17	0,9335
Allow mobility	User modality	Q67	0,6952	-2,6495	0,0823	2,1460	36	83	65	46	0,9355
	User mobility	Q68	–	–	–	–	18	65	165	228	–
	Autonomy	Q69	0,9676	-1,8758	0,1813	2,1046	50	100	81	37	0,9341
Provide navigation	Voice cues	Q70	0,7213	-3,5190	-1,7329	0,9181	21	38	97	75	0,9346
	Portholes/peepholes	Q71	0,8397	-1,3112	0,9292	2,9888	83	121	72	28	0,9339
	Eye-gaze cues	Q72	1,3453	-0,5940	0,6816	1,9304	132	128	85	45	0,9335
	Maps views	Q73	0,6924	-2,1366	-0,0635	2,5259	78	111	131	62	0,9342
	Viewports/Teleports	Q74	1,0107	-0,8492	1,1510	2,7154	97	131	59	29	0,9342
	Objects location	Q75	0,8027	-3,0944	-1,1754	1,3710	30	65	134	84	0,9341

11.2 CASE STUDY RESULTS

We obtained the 820 voluntary participation. We collected 149 observations in the first case study, 422 in the second, and 249 in the last. Table 43 presents the demographic distribution of the mean ability scores obtained (mean θ) for each demographic facet evaluated.

Table 43 – Demographic distribution (global scale)

(a) Age

Obs.	Group	mean θ
9	17 years or less	0.3222
665	18 to 28 years	-0.0229
104	29 to 39 years	0.1875
34	40 to 50 years	-0.0176
8	51 years or more	-0.2625

(b) Gender

Obs.	Group	mean θ
559	Male	-0.0193
241	Female	0.0883
20	Other	-0.3000

(c) Familiarity (awareness)

Obs.	Group	mean θ
154	1 - Novice	-0.3364
281	2 - competent	-0.0402
277	3 - proficient	0.0437
108	4 - expert	0.5138

(d) Familiarity (collaboration)

Obs.	Group	mean θ
72	1 - Novice	-0.0083
218	2 - competent	-0.1504
372	3 - proficient	-0.0677
158	4 - expert	0.3994

(e) Familiarity (environment)

Obs.	Group	mean θ
154	1 - Novice	-0.3364
281	2 - competent	-0.0402
277	3 - proficient	0.0437
108	4 - expert	0.5138

(f) Environmmet

Obs.	Group	mean θ
39	Discord	0.2308
148	Google Docs	0.4284
143	Google Meet	-0.0888
86	Google Sheets	-0.2407
104	Microsoft Teams	0.1202
149	Moodle	-0.3329
71	Moodle (BBB)	-0.1380
15	Skype	0.2067
15	Trello	0.1333
35	Zoom	0.1628
15	Other	0.1067

Regarding gender, we collected 559 male observations (68.17%) and 241 female observations (29.39%); 20 participants did not answer this question or chose the "other gender" option (2.44%). We collected 665 observations from individuals aged 18 to 28 years (81.10%), 104 from individuals aged 29 to 39 years (12.68%), and 34 from individuals between 40 and 50 years old (4.15%). We also collected a minority sample of 8 individuals over 50 years ($0,98\%$) and 9 observations of individuals under 18 (1.10%).

We generated frequency histograms based on the participant's scores and demographic facets (Figures 46 to 51). To verify whether the model presents a distinction in discrimination values (or average) of different groups, we also calculated the normal distribution and the mean score grouped by each demographic perspective. As demographic data, we collected age, gender, preferred 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 Figures 46 to 51, the normal curves generated for each group were significantly close, indicating that our model maintained the behavior already observed in previous scenarios and 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).

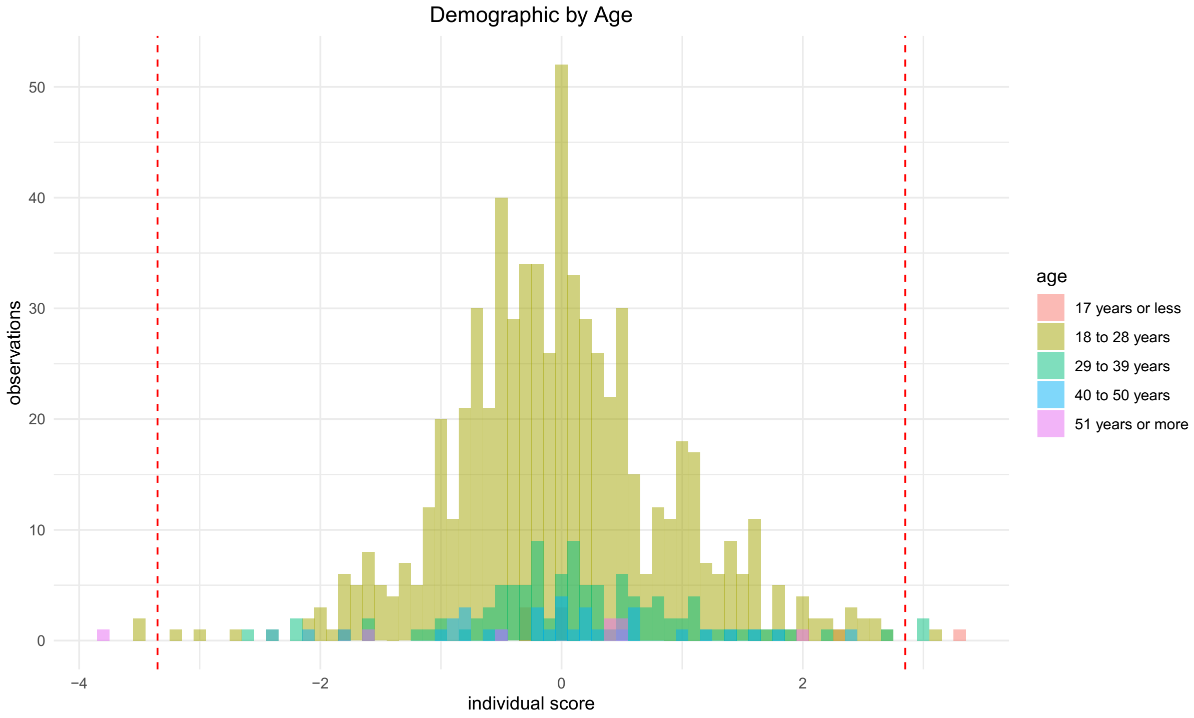
In the demographic facet of participants' age (Figure 46), the normal distribution and the sigmoid function do not present a score distortion in 3 of the 5 age groups evaluated. The group of young individuals, 18 to 29 years old, gave a slight left-shift in the sigmoid function, demonstrating that, generally, older people (29 to 39 years) use these environments more straightforwardly than other age groups. This factor may have positively 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 (Figure 47), we 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. In this global awareness scale scenario, we use a larger set of assessment elements (73 of the 75 assessment items) for the scale parameterization. The model remained representative in all sample groups.

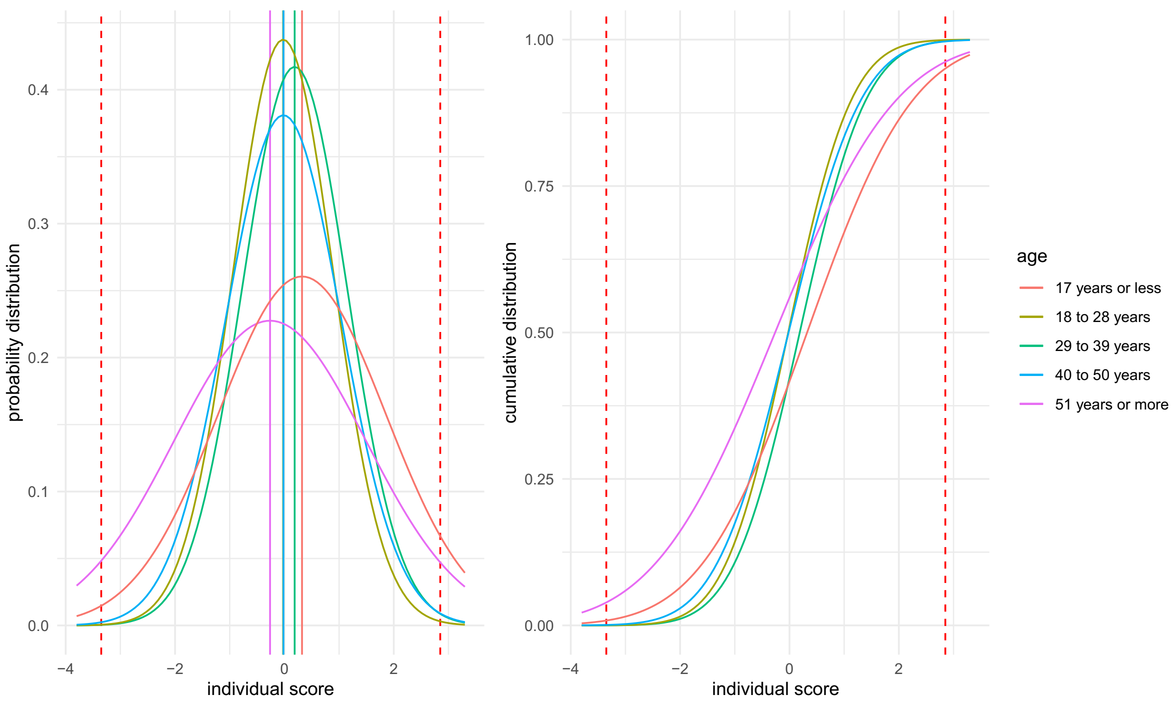
Comparing the scores grouped by the preferred environment (Figure 48), we observed that environments Moodle, Google Sheets, Moodle (BigBlueButton), and Google Meet showed a slight distinction in the average difficulty parameters (namely, mean θ = -0.3329, -0.2407, -0.1380, and -0.0888). On the other hand, in Microsoft Teams, Trello, Zoom, Skype, Discord, and Google Docs environments, sigmoid curves slightly shifted to the right (namely, mean θ = 0.1202, 0.1333, 0.1626, 0.2067, 0.2308, and 0.4284). Positive mean θ 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 (Figures 49a, 50a, and 51a), whether familiarity with the preferred environment, collaboration, or awareness concepts, both normal distribution and probability cumulative distribution (Figures 49b, 50b, and 51b) 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 Figures 46 to 51, the model does not differentiate the scale by a specific group of individuals; the factor that distinguishes individuals is the latent trait evaluated.

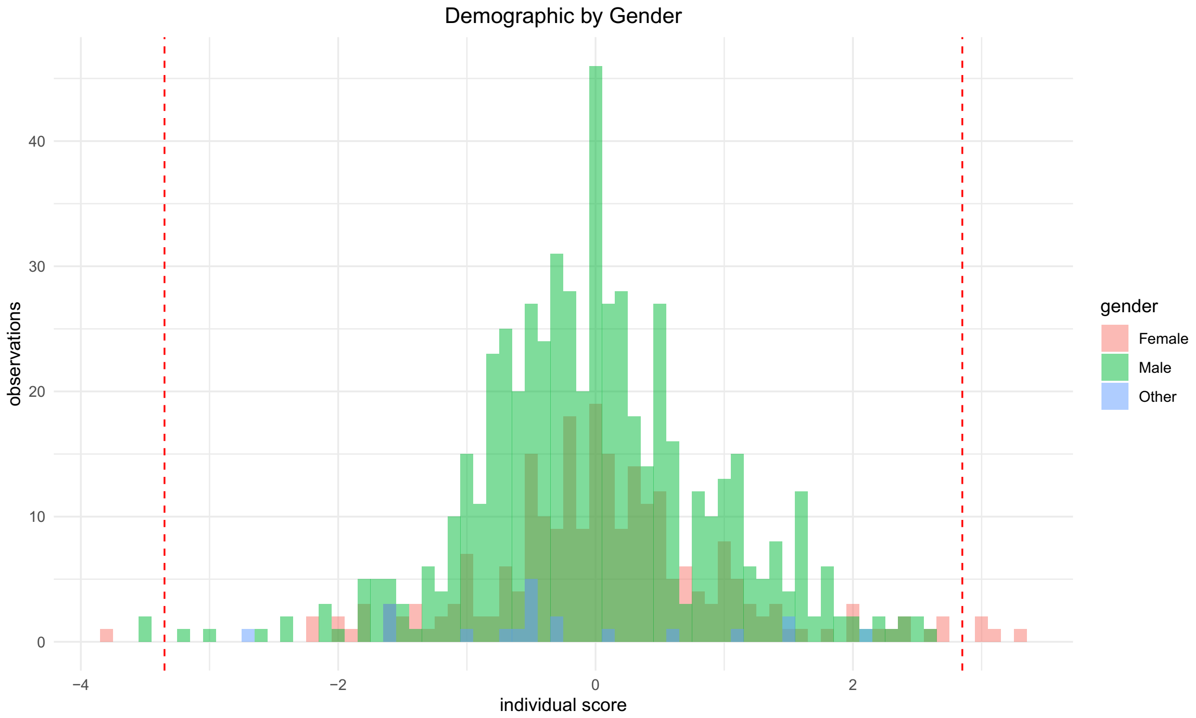


(a) Score

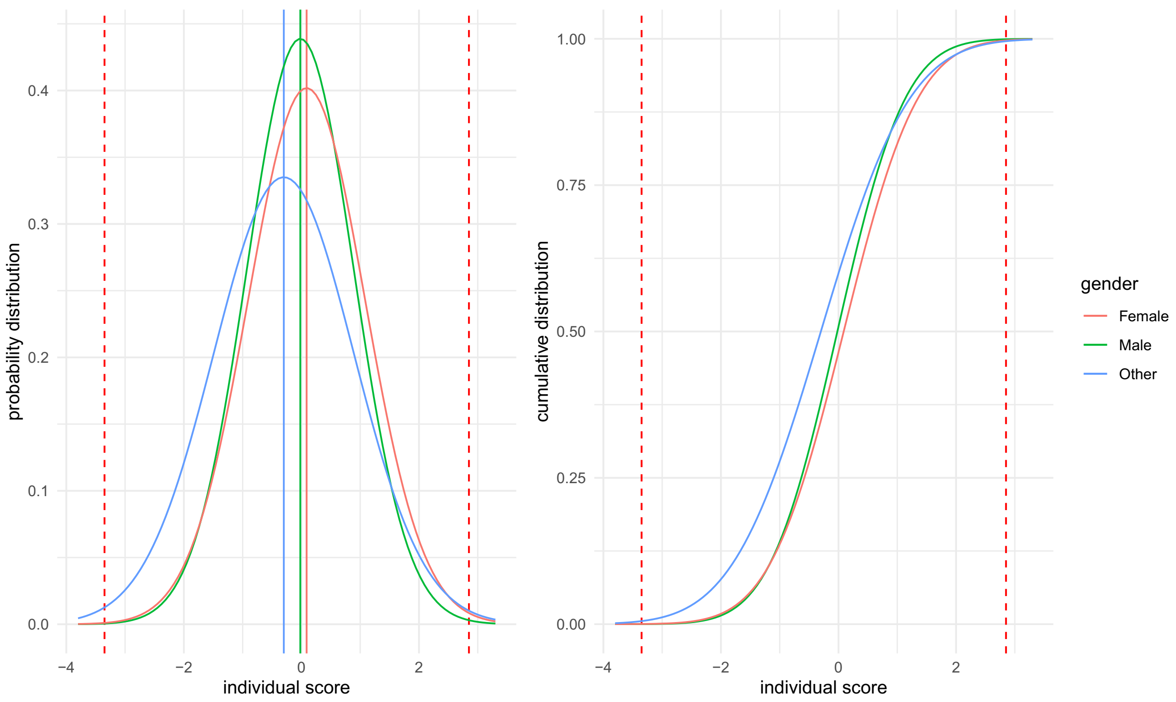


(b) Normal curves

Figure 46 – Demographic distribution of individual score by age (global scale)

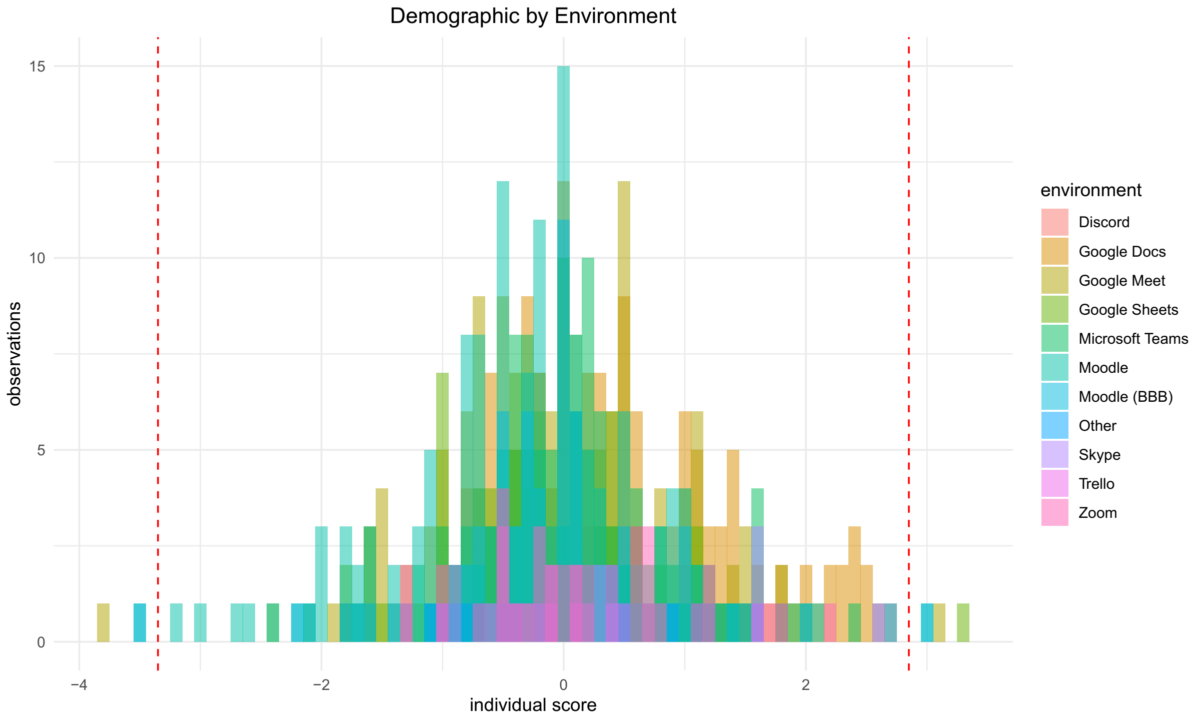


(a) Score

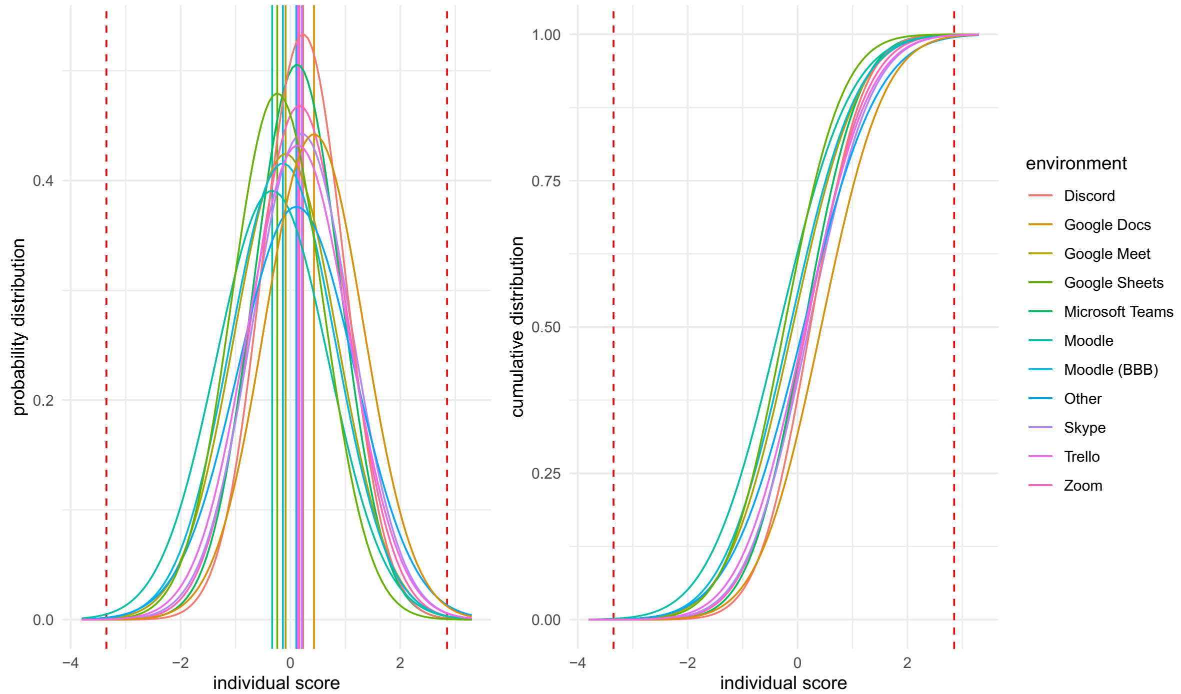


(b) Normal curves

Figure 47 – Demographic distribution of individual score by gender (global scale)

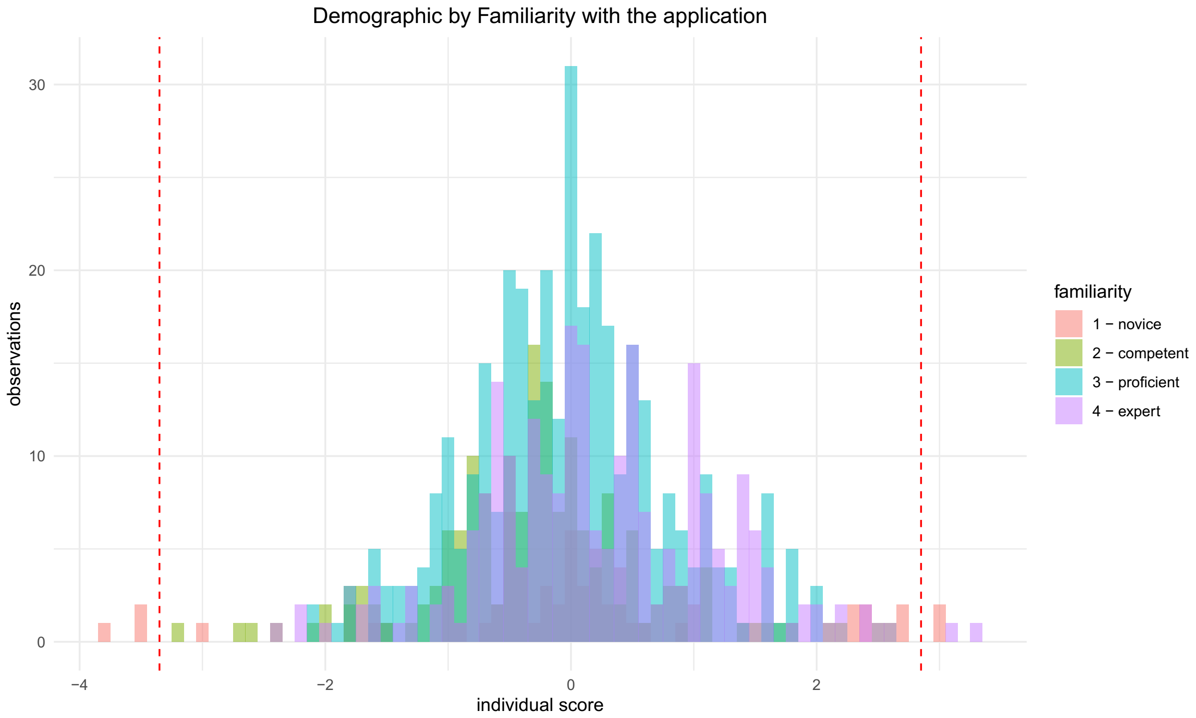


(a) Score

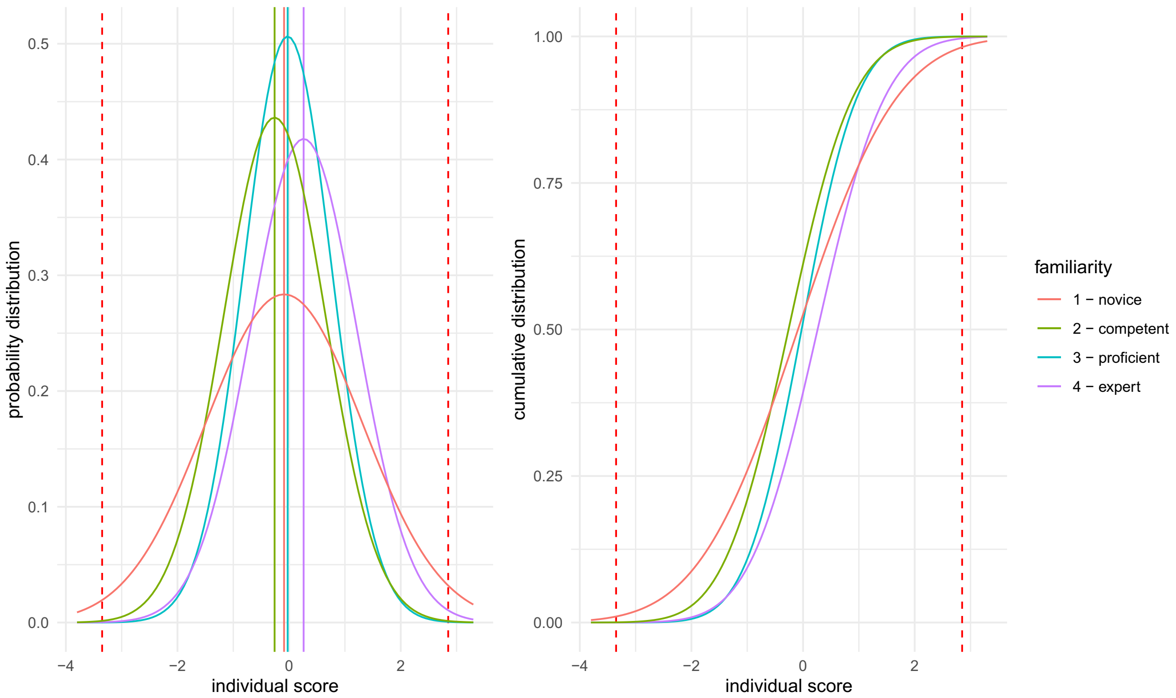


(b) Normal curves

Figure 48 – Demographic distribution of individual score by environment (global scale)

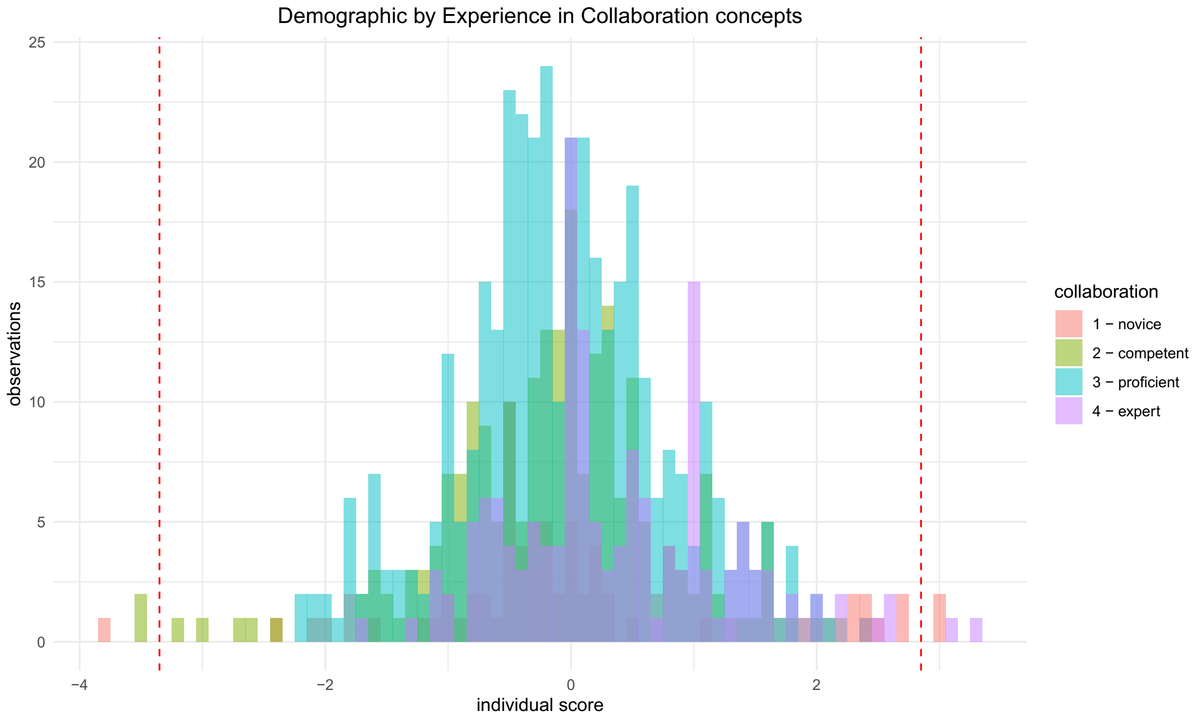


(a) Score

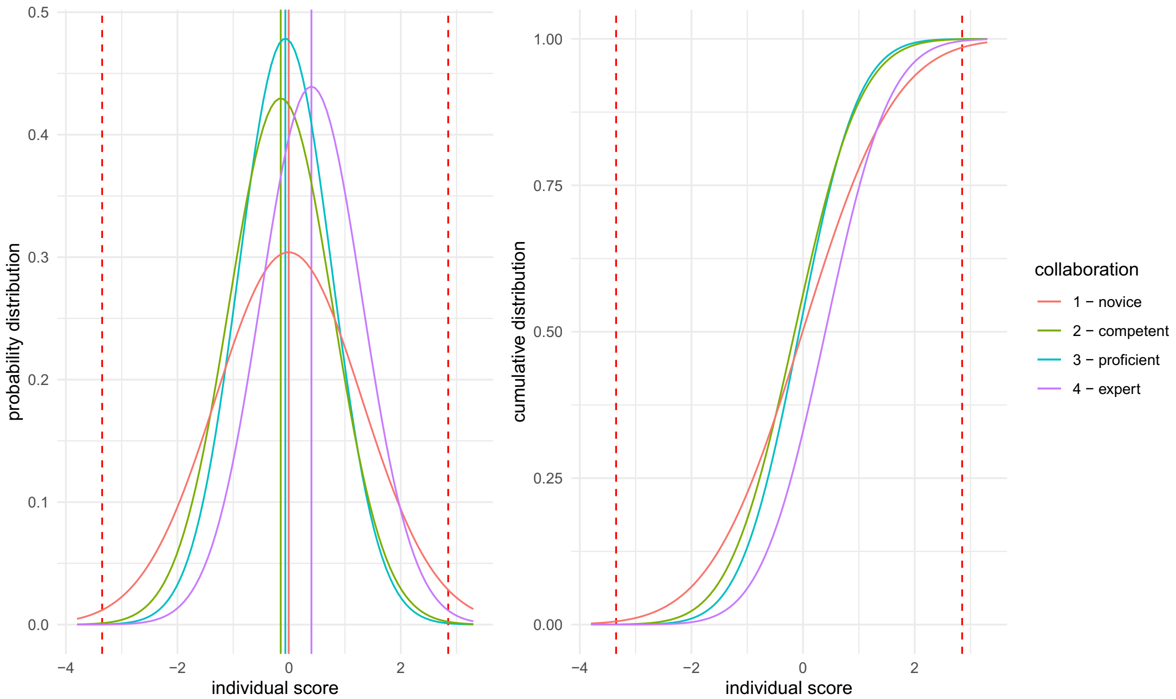


(b) Normal curves

Figure 49 – Demographic distribution of individual score by familiarity (global scale)

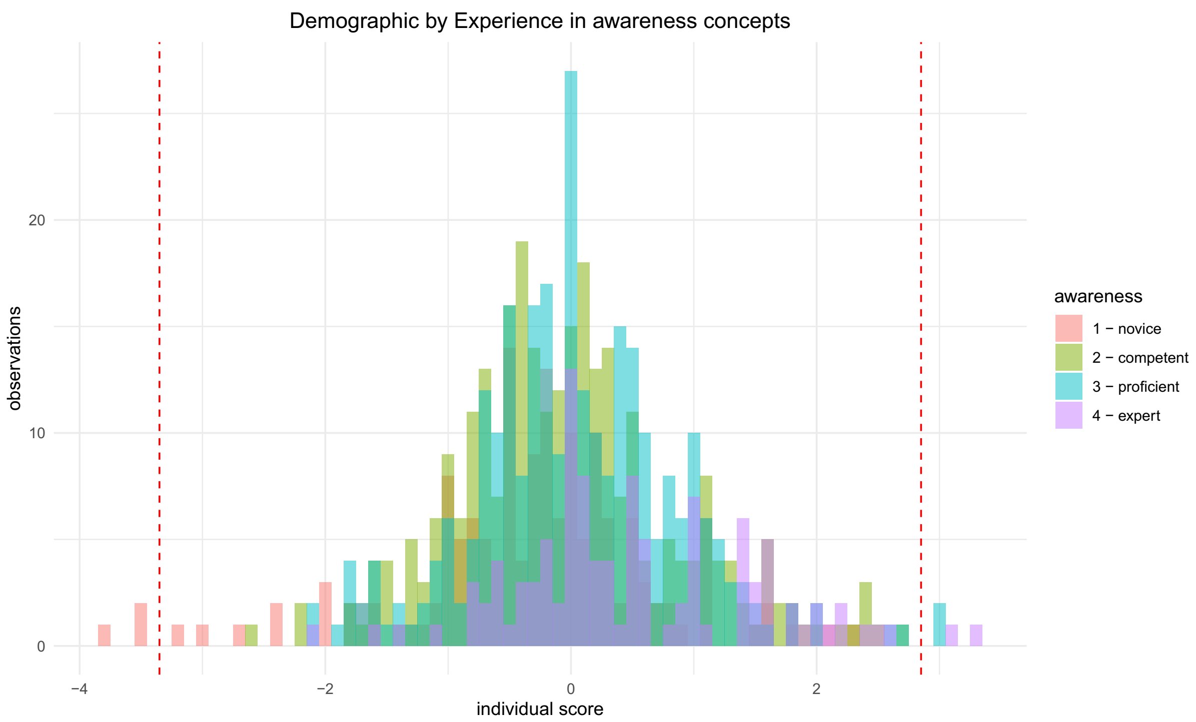


(a) Score

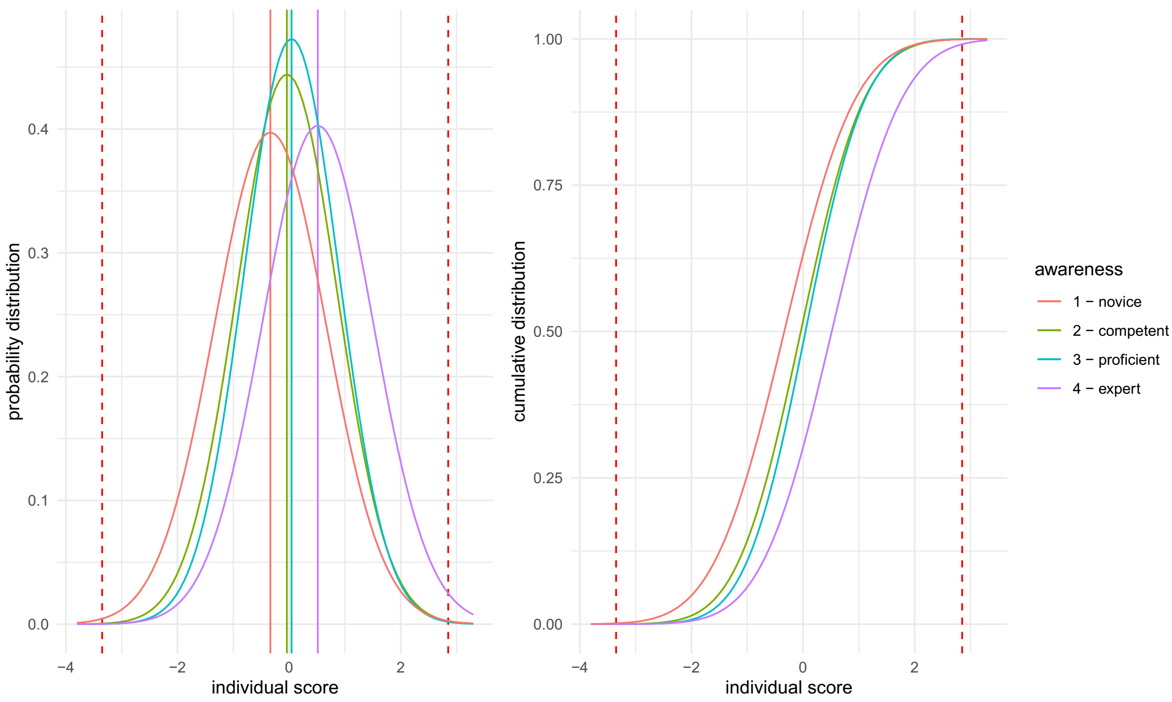


(b) Normal curves

Figure 50 – Demographic distribution of individual score by collaboration (global scale)



(a) Score



(b) Normal curves

Figure 51 – Demographic distribution of individual score by awareness (global scale)

For each awareness mechanism of our calibrated model, we calculated the relationship between the probability of each response item (from strongly disagree to strongly agree) concerning the individual's ability scale. Figure 52 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 (THISSEN; WAINER, 2001); the red dotted line represents the standard error $SE(\theta)$. The intersection point represents the limits at which the model is more representative.

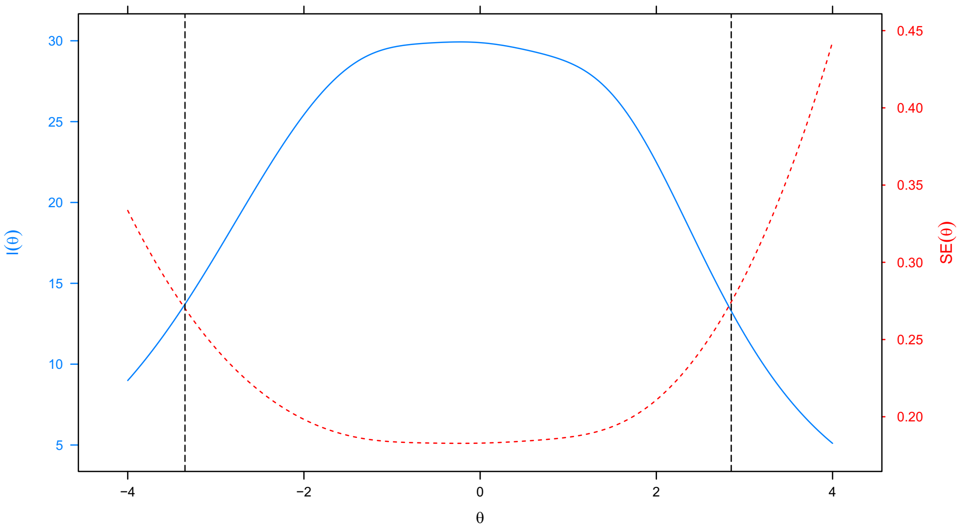


Figure 52 – Test information and SE (global scale)

The Test information and SE curves show the instrument's accuracy. The SE curve reaches its minimum value precisely at the point on the scale where the information curve reaches its maximum ($I(\theta) \approx 30$). 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 $[-3.35, +2.85]$. 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 $(\theta_j > 1)$.

As shown in Figure 52, our model is more representative at the interval $[-3.35, +2.85]$. To cover the outlier scores of individuals with lower or higher abilities ($< 1\%$), we assume in our global awareness support scale a coverage interval $[-4.0, +4.0]$.

11.2.1 The Global Awareness Support Scale

Applying the awareness measurement formulas 15 to 17, we calculated the probability scales $P_{i,k}(\theta_j)$ for the assessment element through our IRT awareness assessment model. To calibrate our model and construct the graphs shown in Figure 53, we considered the scale with a mean μ equal to 0, and a standard deviation σ equal to 1 (ANDRADE; TAVARES; VALLE, 2000).

Figure 53 presents the probability scales generated for each assessment item and awareness dimension. 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.

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.

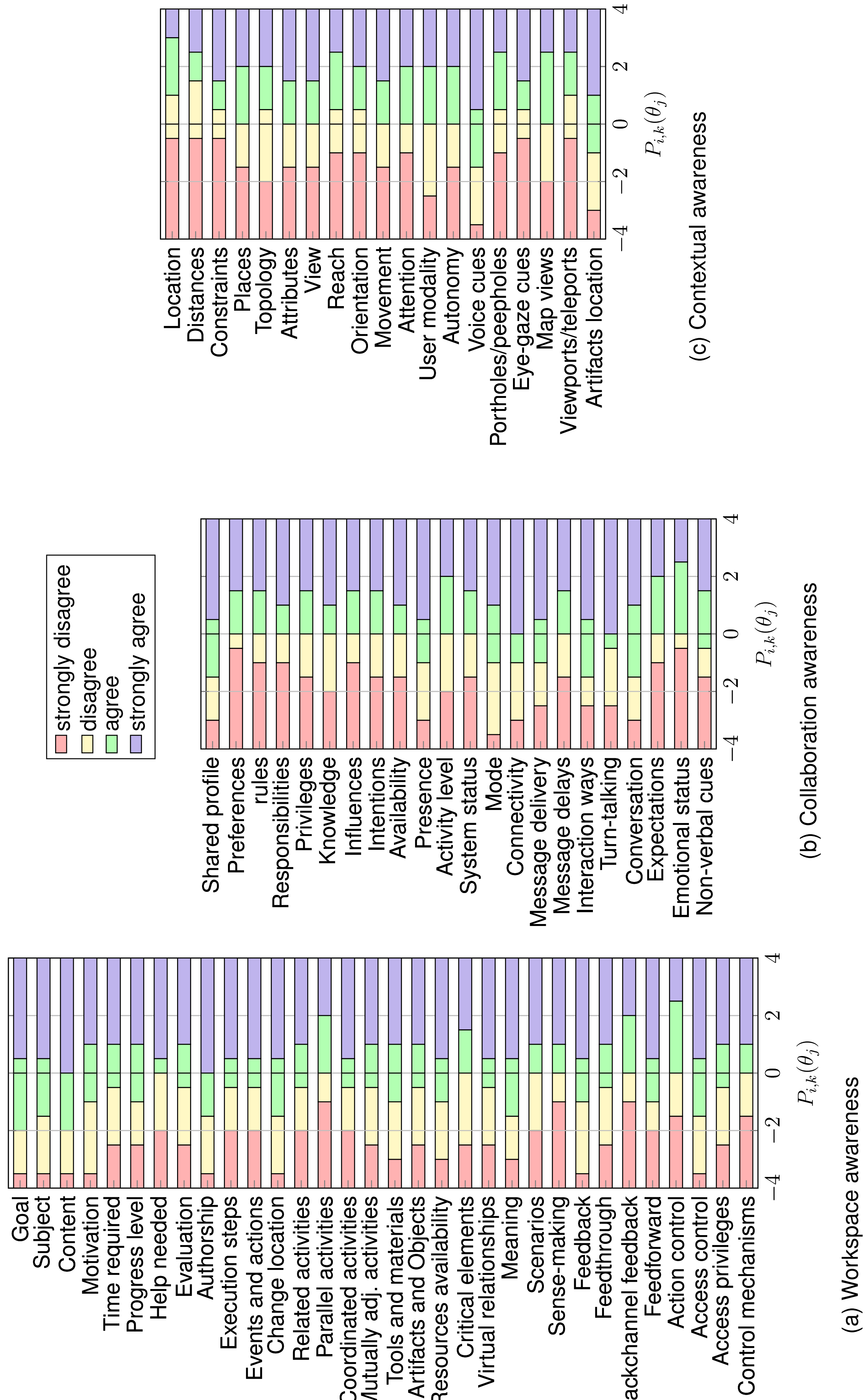


Figure 53 – Ability level scales (global scale)

11.2.1.1 Items positioning

To position the items on the global awareness scale and identify the quality levels, we considered the probability parameter $P_{i,k}(\theta) \geq 0.5$ and the θ and θ^* 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.

To overcome a possible scale misinterpretation, we adopted the principle of invariance of the IRT scales (BAKER, 2001; BAKER; KIM, 2017) and applied a linear transformation (Equation 18) to establish a more appropriate and easier reference for people to interpret their awareness score through a positive scale θ^*. The linear transformation equation is described in Section 8.2.3.

Tables 44 to 46 present, respectively, the workspace, the collaboration, and the contextual awareness quality scales and related assessment items positioning. For the global awareness scale, we consider four levels of awareness support: low $(\theta < -1)$ or $(\theta^* < 90)$; moderate $(-1 \leq \theta \leq 0)$ or $(90 \leq \theta^* \leq 100)$; good $(0 \leq \theta \leq 1)$ or $(100 \leq \theta^* \leq 110)$; and excellent $(\theta > 1)$ or $(\theta^* > 110)$.

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 for both to identify how awareness works and how the collaboration takes place in these environments. Essentially, this model provides reflections toward collaborative improvements by gradually prioritizing supported awareness elements over a participant's perspective. Creating the global awareness scale and its interpretation follows as discussed in the videoconferencing case study results (Section 9.2.2).

Table 44 – Workspace awareness (global scale)

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. Basic workspace awareness information related to the involvement between activities, workflow, and participants, such as goal, subject, content, authorship, change location, and meanings, can be captured by users with a little more skill (θ values closer to -1).
Moderate $(-1 \leq \theta \leq 0)$ $(90 \leq \theta^* \leq 100)$	The collaborative environment rarely provides workspace capabilities or information about the activities, environment, or workflow. It provides a moderate interaction or understanding of artifacts and objects shared in the workspace. Due to the workspace limitations, the interaction is difficult. In addition to the basic workspace awareness information partially identified in the previous level, the participants understand much of the workspace mechanisms in some way. On the other hand, more complex workspace awareness information about workflow, interaction, understanding, and relationship categories, such as help needed, parallel activities, critical elements, virtual relationships, sense-making, action control, access privileges, and control mechanisms, are difficult to acquire.
Good $(0 \leq \theta \leq 1)$ $(100 \leq \theta^* \leq 110)$	The collaborative environment provide a good support over the workspace capabilities and presents some information like activities, environment, and workflow. The interaction and understanding of artifacts and objects shared is possible, although it usually does not present full operability. At this level, all workspace awareness mechanisms can be accessed by participants properly. Users fully reach the most common awareness aspects with a little more skill (θ values closer to 1).
Excellent $(\theta > 1)$ $(\theta^* > 110)$	The collaborative environment fully 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 45 – Collaboration awareness (global scale)

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. Users with more skill can capture basic identity, status, and communication facets, like identity mechanism, shared profile, presence, mode, connectivity, message delivery, interaction ways, and conversation (θ values closer to -1).
Moderate $(-1 \leq \theta \leq 0)$ $(90 \leq \theta^* \leq 100)$	The collaborative environment presents a moderate social interaction and collaboration aspects over the same mechanisms as perceived in the previous level. The environment provides some 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 in the collaboration process. The closer the value of θ gets to 0, the better participants take advantage of the collaboration aspects.
Good $(0 \leq \theta \leq 1)$ $(100 \leq \theta^* \leq 110)$	The collaborative environment presents good social interaction and collaboration aspects. The environment provides adequate status, identity, and the participant's capabilities. The awareness information is relevant to the participant's interests, and they recognize that the content helps the collaboration process. At this level, all collaboration awareness mechanisms can be accessed by participants properly; Users fully reach the most common collaboration aspects with a little more skill (θ values closer to 1).
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 46 – Contextual awareness (global scale)

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. Understanding the contextual mechanisms depends basically on the individual effort; the adequate appropriation of contextual support elements by participants over the collaborative interface is unexpected.
Moderate $(-1 \leq \theta \leq 0)$ $(90 \leq \theta^* \leq 100)$	The collaborative environment provides moderated access to contextual information. Closer θ value to 0, participants partially reach environmental spatiality and navigation; however, members' mobility remains not archived. Although some contextual information can be perceptible to participants with this ability level, providing a context-sensitive collaborative environment remains challenging. The interaction process does not consider contextual aspects, or these contextual elements are unnecessary or unrequired for a certain level of collaborative work.
Good $(0 \leq \theta \leq 1)$ $(100 \leq \theta^* \leq 110)$	The collaborative environment provides good access to contextual information while the participant's ability approaches $\theta = 1$. Participants reach environmental navigation and spatiality aspects, and the environment provides some operability over participants' contextual information. At this level, all contextual awareness mechanisms can be accessed by participants properly.
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.

11.2.1.2 Scale interpretation

In the first global awareness scale perspective, as shown in Figure 53, 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), we represent the expected ability intervals in which participants present a certain probability P_i of selecting each response category presented. In other words, starting from the participants' ability scale θ, we represent the probable intervals $P_i(\theta)$ that participants are most likely to correctly identify/understand the awareness mechanism in the evaluated interface.

In the second awareness scale perspective, presented in Tables 44 to 46, we categorized the results concerning the skill levels of the expected participants. In our assessment model, the ability scale θ encompasses within the interval $[-4.0, +4.0]$ and the adjusted ability scale θ^* encompasses within the interval $[+60, +140]$ (awareness points). According to the model calibration results from this global perspective, the model's coverage range is between $\theta = [-3.35, +2.85]$; thus, it is expected that the model adequately represents participants with skill in this range (awareness scale $\theta^* = [77.5, +128.5]$).

As shown in the scales of the previous scenarios (Chapters 8 to 10), in our workspace, collaboration, and awareness global 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 acquire the more important awareness mechanisms easily.

11.3 MODEL RELIABILITY

Similarly to previous scenarios (Chapters 8 to 10), we use the Cronbach's alpha coefficient (α) (DEVELLIS, 2016) to check internal consistency. 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. We also use the IRT technical properties of discrimination (a) and difficulty (b), combined with α coefficient, to assess the Awareness Assessment Model Instrument reliability and internal consistency.

11.3.1 Reliability results

Both alpha and IRT params strongly demonstrate the validation of our proposed model. First, the adequate representation of the awareness scale θ (interval $[-3.35, +2.85]$ as presented in Figure 52) is good evidence of the instrument's reliability.

In addition, the internal reliability through Cronbach's alpha coefficient (DEVELLIS, 2016) demonstrated an excellent internal consistency for all assessment items $(\alpha > 0.93)$.

Second, we calculate the instrument's reliability function $r_{xx}(\theta)$ (THISSEN; WAINER, 2001; CHALMERS, 2012), across participants' latent trail (see Figure 54). Our model shows excellent reliability, and the function reaches its highest value $(r_{xx} > 0.95)$ over the scale region where the information function is representative.

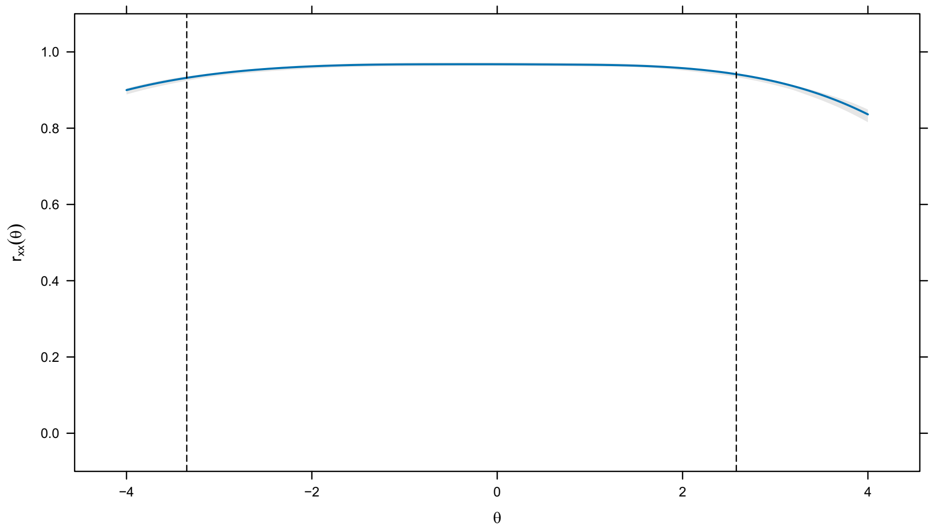


Figure 54 – Test reliability (global scale)

11.4 MODEL DIMENSIONALITY

We used the 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 explain the correlations pattern (OSBORNE, 2014). 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 (THOMPSON, 2004; BROWN, 2015).

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 (HAIR et al., 2009). Factor analysis is based on the simulation of random data to determine the number of factors (IZQUIERDO; OLEA; ABAD, 2014). The factors/components retained with real data must have an eigenvalue higher than those obtained randomly.

We used the same data set (820 observations) for the EFA and CFA. To determine the number of factors retained in the EFA, we will use the Latent Root (or Kaiser) Criterion (HAIR et al., 2009). In the Latent Root, the factors or components retained in the analysis with real data must have an eigenvalue higher than ones obtained randomly (LEDESMA; VALERO-MORA, 2019); thus, only factors with eigenvalues greater than or equal to 1 are considered.

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 (PASQUALI; PRIMI, 2003). 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 (PASQUALI; PRIMI, 2003). Combining EFA and CFA 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 (HAIR et al., 2009).

11.4.1 EFA results

The latent root criterion suggests a strong principal component and a secondary prominent component. As we can see in Figures 55a to 55c, by applying the Kaiser criterion (eigenvalue ≥ 1), we identify two principal components for the workspace, collaboration, and contextual awareness categories.

Very similar to the results of the assessment videoconferencing scenario (Chater 9), the inclination angles decreased sharply from the second factor onwards, approaching the horizontal line of value 1 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.

The EFA results from this global scenario highlight the identification of two principal components in each of the awareness perspectives, which in the previous scenario, respectively, were 2 factors for workspace, 3 for collaboration, and 3 for contextual awareness. We conjecture that this result is due to larger sampling and better calibration of the IRT parameters a, b ($b1$, $b2$, and $b3$), e θ.

The eigenvalue represents the total variance a given principal component can explain. Starting from the first component, each subsequent component is obtained from partially out the previous component. Therefore, the first component explains the most variance, and the last component explains the least. Tables 47 to 49 present the total variance explained by each principal component (pc).

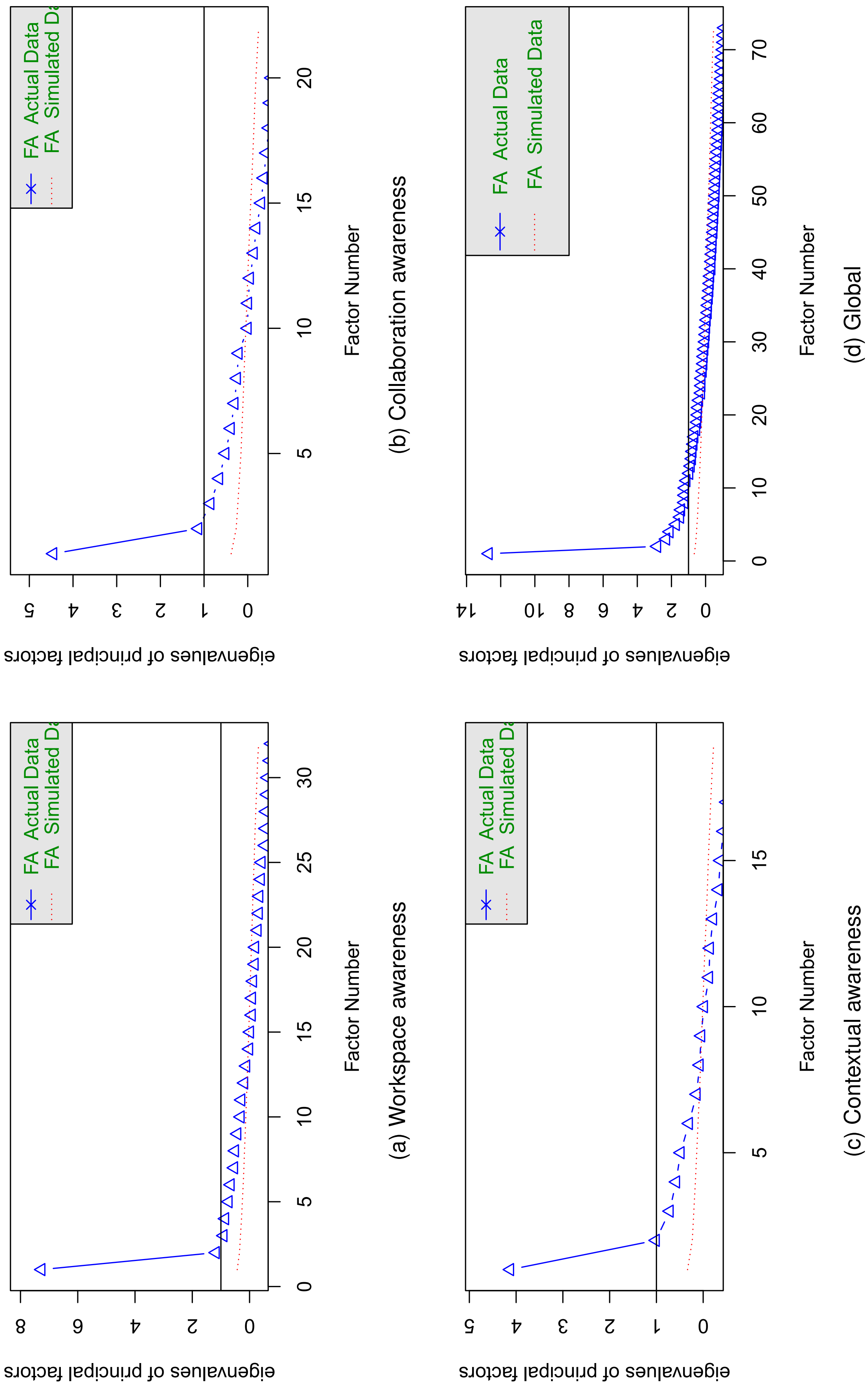


Figure 55 – Model dimensionality (global scale)

Table 47 – Total variance explained: workspace awareness (global scale)

pc	Total	Variance	Cumulative		pc	Total	Variance	Cumulative
1	8.0162	25.05%	25.05%		17	0.7055	2.20%	84.50%
2	1.9745	6.17%	31.22%		18	0.6408	2.00%	86.50%
3	1.7105	5.35%	36.57%		19	0.6195	1.94%	88.44%
4	1.6400	5.12%	41.69%		20	0.5550	1.73%	90.17%
5	1.5194	4.75%	46.44%		21	0.5028	1.57%	91.74%
6	1.4353	4.49%	50.92%		22	0.4697	1.47%	93.21%
7	1.3295	4.15%	55.08%	...	23	0.4342	1.36%	94.57%
8	1.2725	3.98%	59.06%		24	0.3854	1.20%	95.77%
9	1.1987	3.75%	62.80%		25	0.3445	1.08%	96.85%
10	1.0927	3.41%	66.22%		26	0.2324	0.73%	97.57%
11	1.0267	3.21%	69.42%		27	0.2033	0.64%	98.21%
12	0.9371	2.93%	72.35%		28	0.1968	0.62%	98.83%
13	0.9145	2.86%	75.21%		29	0.1655	0.52%	99.34%
14	0.8044	2.51%	77.72%		30	0.1403	0.44%	99.78%
15	0.7461	2.33%	80.06%		31	0.0718	0.22%	100.01%
16	0.7163	2.24%	82.29%		32	-0.0017	-0.01%	100.00%

Table 48 – Total variance explained: collaboration awareness (global scale)

pc	Total	Variance	Cumulative		pc	Total	Variance	Cumulative
1	5.2134	23.70%	23.70%		12	0.73635	3.35%	83.45%
2	1.9408	8.82%	32.52%		13	0.63064	2.87%	86.31%
3	1.6767	7.62%	40.14%		14	0.54007	2.45%	88.77%
4	1.4756	6.71%	46.85%		15	0.51157	2.33%	91.09%
5	1.3336	6.06%	52.91%	...	16	0.46361	2.11%	93.20%
6	1.1972	5.44%	58.35%		17	0.36152	1.64%	94.84%
7	1.1052	5.02%	63.37%		18	0.34154	1.55%	96.40%
8	1.0725	4.87%	68.25%		19	0.27794	1.26%	97.66%
9	0.9903	4.50%	72.75%		20	0.25190	1.14%	98.80%
10	0.8463	3.85%	76.60%		21	0.15232	0.69%	99.50%
11	0.7701	3.50%	80.10%		22	0.11088	0.50%	100.00%

In the workspace awareness perspective, Factor 1 explains 25.05% ($pc = 8.0162$); in the contextual awareness perspective, Factor 1 explains 27.44% ($pc = 5.2134$); and in the collaboration awareness perspective, Factor 1 explains 22.17% ($pc = 4.8774$) 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 (RECKASE, 1979).

Table 49 – Total variance explained: contextual awareness (global scale)

pc	Total	Variance	Cumulative		pc	Total	Variance	Cumulative
1	4.8774	25.67%	25.67%		11	0.7165	3.77%	84.48%
2	1.7716	9.32%	34.99%		12	0.6231	3.28%	87.76%
3	1.5511	8.16%	43.16%		13	0.5627	2.96%	90.72%
4	1.4037	7.39%	50.55%		14	0.4762	2.51%	93.22%
5	1.2806	6.74%	57.29%	...	15	0.3694	1.94%	95.17%
6	1.0875	5.72%	63.01%		16	0.3078	1.62%	96.79%
7	0.9458	4.98%	67.99%		17	0.2781	1.46%	98.25%
8	0.8352	4.40%	72.38%		18	0.1832	0.96%	99.22%
9	0.8226	4.33%	76.71%		19	0.1490	0.78%	100.00%
10	0.7584	3.99%	80.70%					

11.4.2 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 (HAIR et al., 2009). Based on the EFA results, we rewrote the awareness perspectives based on two factors for each awareness perspectives.

Figures 56, 57, and 42 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 (PASQUALI; PRIMI, 2003; TAVAKOL; WETZEL, 2020). Extracting the evidence from the factor analysis of the model presented in Figures 56, 57, and 42, we identified the main latent dimensions (factors) for the workspace and two for collaboration and contextual awareness perspective, as shown in Tables 50, 51, and 52.

We could visualize a significant equivalence by comparing the model generated by the EFA analysis of the assessment items of our taxonomy (awareness elements and design categories). Furthermore, as shown in Figures 56, 57, and 42, most of the factor loadings exceed 0.3. The confirmatory factor analysis results largely maintained the structure defined in our taxonomy, and the factor loadings of the CFA items demonstrate the instrument's construct validity (PASQUALI; PRIMI, 2003).

Finally, we calculated Composite Reliability (CR) to evaluate the construct validity of the proposed model (HAIR et al., 2009). The measurements obtained were then evaluated following the recommendations of Fornell and Larcker (1981). All factors in the model must present a CR value above 0.7 to demonstrate the instrument's construct validity (FORNELL; LARCKER, 1981; HAIR et al., 2009).

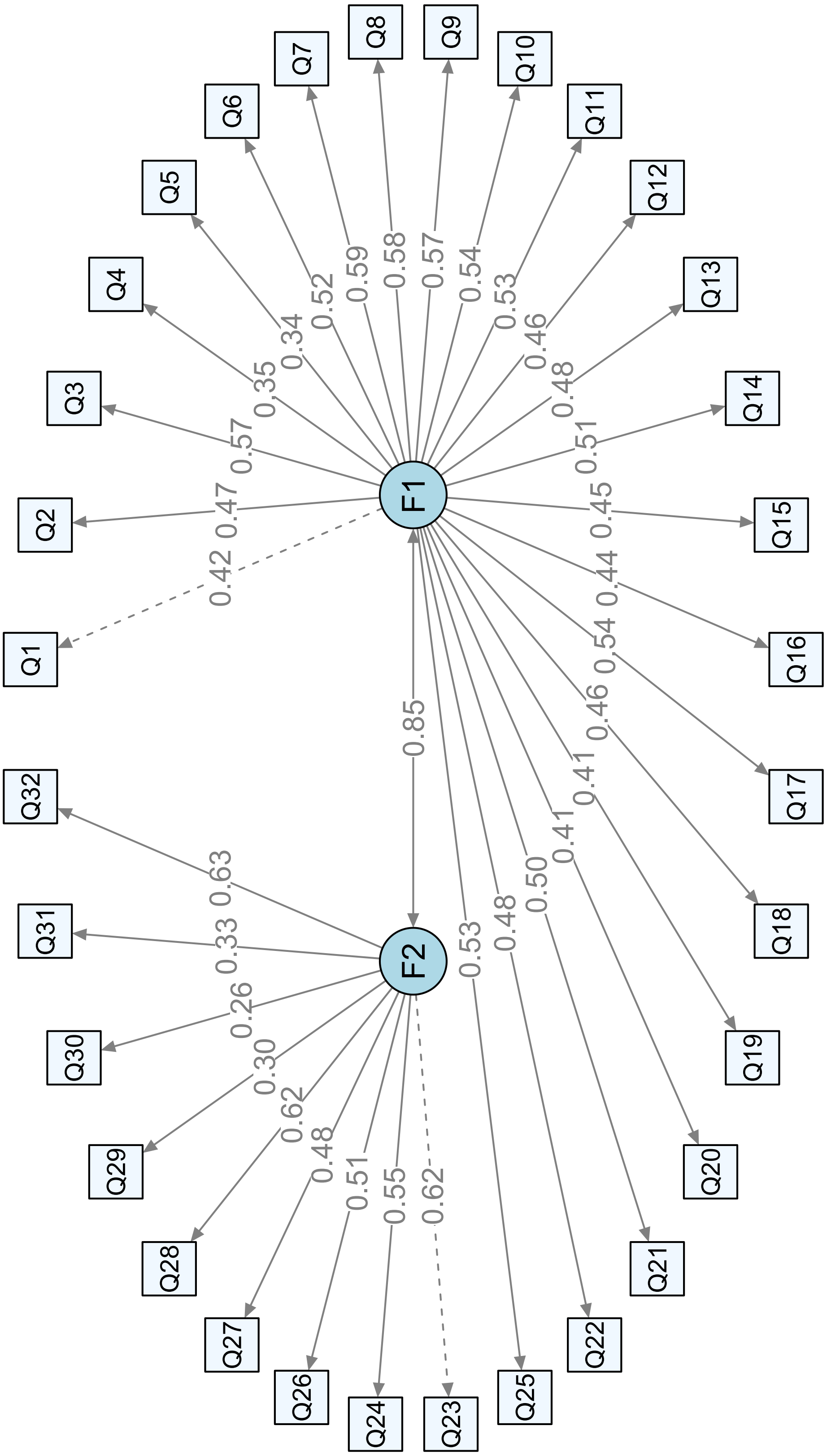


Figure 56 – Workspace awareness (global scale)

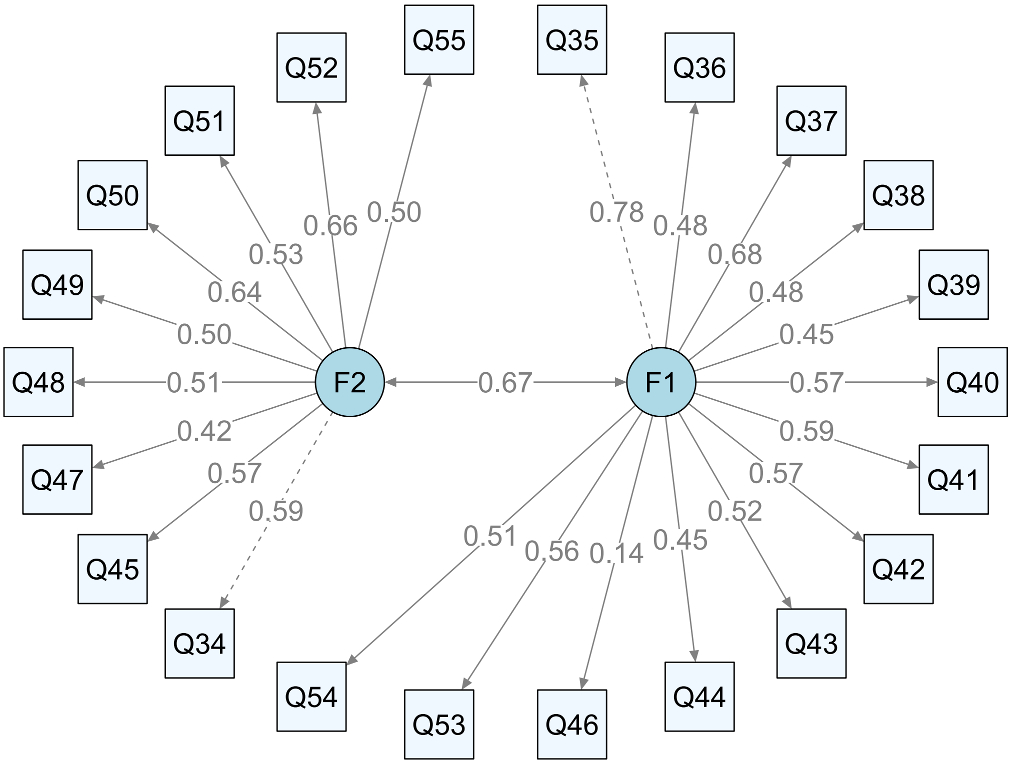


Figure 57 – Collaboration awareness (global scale)

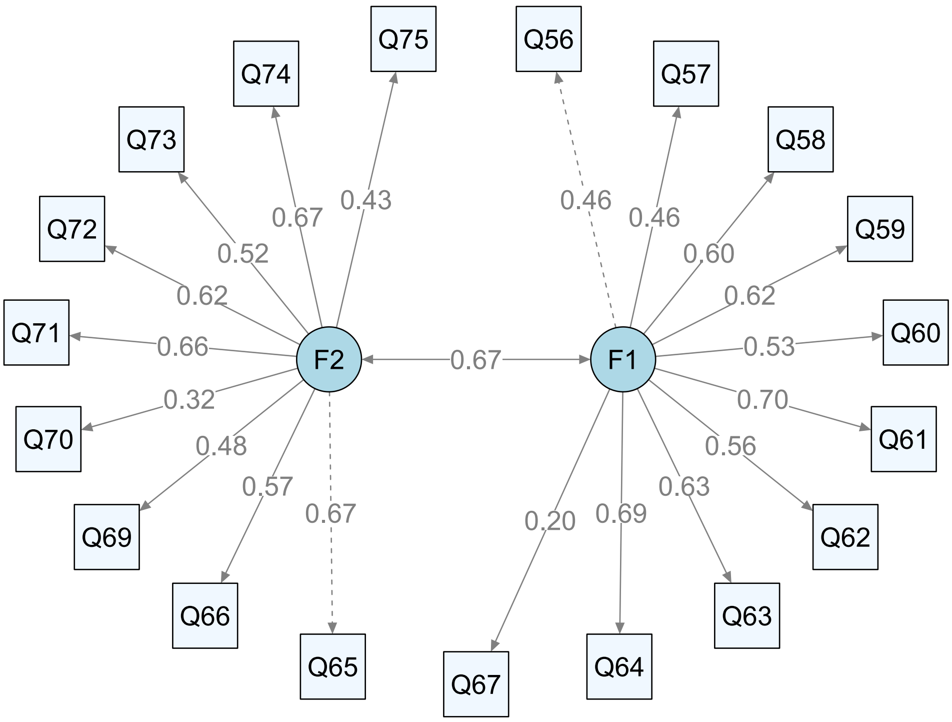


Figure 58 – Contextual awareness (global scale)

Evidence of construct validity indicates that the items measured in the sample represent the real measurements in the population (HAIR et al., 2009). In the workspace awareness perspective, factors F1 and F2 presented CR values of 0.870 and 0.732, respectively. In the collaboration awareness perspective, factors F1 and F2 presented CR values of 0.834 and 0.795, respectively. In the contextual awareness perspective, values of 0.801 and 0.791 were found for factors F1 and F2.

Table 50 – Factor analysis of the workspace awareness perspective (global scale)

Factor	Factor description
F1 (Factor 1)	**The environmental understanding of the shared workspace**. It provides an understanding of the collaborative activities (Q1 to Q8 – goal, subject, content, motivation, time required, progress level, help needed, and evaluation), involves considerations regarding the workflow (Q9 to Q16 – authorship, execution steps, events and actions, change location, related activities, coordinated activities, and adjusted activities) and environment (Q17 to Q21 – tools and materials, artifacts and objects, resources availability; It also related to support meanings and feedback mechanisms (Q22 and Q25).
F2 (Factor 2)	**The participant's feedback over the shared workspace**. It involves the understanding of the scenarios and sense-making cues (Q23 and Q24) and the support of the interaction (Q26 to Q28 – feedthrough, backchannel feedback, and feedforward) and control mechanisms relationship (Q29 to Q32 – action control, access control, assess privileges, and control mechanisms).

Table 51 – Factor analysis of the collaboration awareness perspective (global scale)

Factor	Factor description
F1 (Factor 1)	**The participant's capabilities to collaborate**. It is related to supporting the participant's preferences (Q35) and collaboration capabilities (Q36 to Q41 – rules, responsibilities, privileges, knowledge, influences, and intentions). It also provides some status (Q42 to Q44 – availability, presence, and activity level), communication mode (Q46), and social interaction facilities (Q53 and Q54 – expectations and emotional status).
F2 (Factor 2)	**The participants' social communication**. It refers to the participants' shared profile (Q34) and status (Q45) support, combined with resources to establish communication (Q47 to Q52 – connectivity, message delivery, message delays, interaction ways, turn-talking, and conversation) and non-verbal cues considerations (Q55).

Table 52 – Factor analysis of the contextual awareness perspective (global scale)

Factor	Factor description
F1 (Factor 1)	**The virtual setting of the contextual environment**. It involves considerations about contextual spatiality (Q56 to Q64 – location, distances, restrictions, places, topology, attributes, view, reach, and orientation) and user modality (Q67).
F2 (Factor 2)	**The common information about the shared environment navigation**. It involves some spatiality cues (Q65 and Q66 – movement and range of attention) combined with contextual autonomy (Q69) and navigation support (Q70 to Q75 – voice cues, portholes/peepholes, eye-gaze cues, map views, viewports/teleports, and artifacts location).

11.5 DISCUSSION

To calibrate the IRT model and generate the global awareness scale, we used a set of 820 observations, considering the three scenarios from previous case studies. In the first two scenarios (571 observations), we used the full version of the data collection instrument, consisting of 10 questionnaires (test books) provided in the full version of the model. In the last scenario, we exposed the instrument to different groups of evaluators, where each team developed its own set of evaluation items (questionnaire) and prepared the evaluation in a specific scenario. As we also use this collected data (249 observations) to generate the global scale, our sample is now partially balanced by the BIB technique (HINKELMANN; KEMPTHORNE, 2005).

In the global awareness scale results, we found suitable indicators from the perspective of demographic data and IRT parameterization. Then, the participant's ability scores and awareness quality scales were constructed based on the 73 calibrated items. Only two awareness mechanisms did not have a satisfactory calibration, and these were disregarded from the final version of the scale due to their low discrimination on the ability scale, namely Q33 (identity) and Q68 (user mobility).

We believe that assessment item Q33 did not provide adequate calibration results because providing participants' identity is an elementary awareness aspect in a collaborative scenario; so, most participants positively indicated their understanding of this mechanism, resulting in an inconsistent calibration for the discrimination and difficulty parameters. As for the second item, the clearest conjecture is that allowing user mobility is somewhat difficult to evaluate in a short time window, as it does not depend exclusively on the latent trait evaluated (support for awareness). It is related to the specific peculiarities of the context and target device available to access the collaborative environment. An alternative scenario would be to use the target environment for a longer period and in a context that encourages the exploration of this mobility facet.

In general, the results of the global scale assessment were positive, and the most familiar environments presented the best performance. Discord, Google Docs, Microsoft Teams, Skype, Trello, and Zoom presented a better (positive) overall awareness score, respectively, θ equals to 0.2308, 0.4284, 0.1202, 0.2067, 0.1333, and 0.1628; the adjusted ability scores θ^* were equal to 102.3, 104.3, 101.2, 102.1, $101,4$, and 101.7 awareness points. Users of Google Meet, Google Sheets, Moodle, and Moolde (BBB) a slightly greater difficulty in identifying awareness information, respectively, θ equals to -0.0888, -0.2407, -0.3329, and -0.1380; the adjusted ability scores θ^* were equal to 99.1, 97.5, 96.6, and 98.6 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 greater ability scores can identify properly existing awareness mechanisms.

To construct the ability scale θ, we calculated the probability $P_{i,k}(\theta_j)$ considering the Samejima's gradual scale (SAMEJIMA, 1969). The generated awareness support scale presented a coverage interval $[-4.0, +4.0]$, with the most appropriate values in interval $[-3.35, +2.85]$. 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 global awareness scale for assessing awareness using Item Response Theory. Our global scale presents gradual ability scale levels of the latent trait that make it possible to interpret the degree of skill an evaluator has given that he or she used the proposed measurement instrument.

The global scale validation was very positive; we exposed the assessment model in different scenarios to assess the model's internal consistency, reliability, and dimensionality. In all 73 calibrated assessment items, θ, 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 (> 0.95).

We used the exploratory factor analysis (EFA) and the confirmatory factor analysis (CFA) to test the model dimensionality. The EFA results indicated a strong tendency towards the one-dimensional model (latent root criterion) (HAIR et al., 2009), legitimating the correlation between the assessment items and the observed latent trail. The CFA results demonstrate the instrument construct validity: all factors presented adequate composite reliability (CR > 0.7) (FORNELL; LARCKER, 1981; HAIR et al., 2009), and most of the factor loadings are suitable (≥ 0.3) (TAVAKOL; WETZEL, 2020; PASQUALI; PRIMI, 2003).

12 CONCLUSION AND FUTURE WORK

Awareness and collaboration concepts are intrinsically related, and their understanding expanded in the same way as research in the field evolved. We observed efforts to establish common sense about what awareness is, what it represents, and what it is related to. However, achieving an accurate and clear definition of awareness remains a challenge.

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. Providing adequate awareness mechanisms ensures support for the collaboration process, consolidating awareness as the cornerstone of collaborative environments.

We believe awareness intrinsically relates to participants' skills to identify, understand, and project their actions. Thus, properly assessing a collaborative environment support is possible if we consider the awareness elements from the participant's perspective. As the participants' understanding differs, the awareness support scale must represent individuals with lower or higher abilities.

In this work, we apply 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, our model seeks to build a more faithful representation of these existing profiles across a broad spectrum of individual abilities.

12.1 METHOD SYNTHESIS

We conducted rigorous methodological procedures to establish and validate the awareness assessment model.

First, we conduct a systematic mapping study on identifying awareness support in the context of the collaborative system (Chapter 3). We followed the guidelines in Petticrew and Roberts (2006), Kitchenham and Charters (2007) and Petersen, Vakkalanka, and Kuzniarz (2015). Regarding awareness support, we identified a set of 92 supporting awareness elements or widgets organized into 17 design categories that address five main awareness dimensions. It represents a set of awareness elements synthesized concerning the consolidated classification of the 5W+1H framework (GUTWIN; GREENBERG, 2002).

Second, based on the systematic mapping results, we elaborate an awareness taxonomy that considers the awareness elements and collaboration aspects necessary for cooperative work (Chapter 4). Our Taxonomy Definition Method is based on Bailey's conceptual approach (BAILEY, 1994) and combines the guidelines presented by Nickerson, Varshney, and Muntermann (2013), Usman et al. (2017), and Szopinski, Schoormann, and Kundisch (2019). As a result, we built a multidimensional taxonomy represented in three main awareness dimensions, namely, collaboration, workspace, and contextual, encompassing 75 awareness mechanisms described in the CSCW literature.

Third, 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 (Chapter 5). 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 believe that the essential aspects of the collaboration process are provided through adequate support for each view of awareness. 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 (MANTAU; BENITTI, 2023)). To use the proposed assessment method properly, we designed an assessment process inspired by the HCI guidelines and the recommendations for evaluating software product quality of ISO/IEC 25040:2011 (STANDARDIZATION, 2011). In this way, we designed an adaptive approach, 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.

We validated the assessment model from expert panel and case study perspectives. First, to improve the proposed model, we expose the model's artifacts to expert appreciation through the expert panel approach (Chapter 7) (BEECHAM et al., 2005). In this stage, we exposed the model's artifacts to HCI and collaborative systems expert scrutiny to collect an accurate model's criterion and content validity by analyzing the measurement instrument items' usefulness aspects (clarity, relevance, consistency, and completeness). The results suggest practical applicability and a clear definition of model artifacts and their structure; the artifacts of the assessment model (the questions or items) are representative, clear, and relevant, allowing them to be adopted in different situations. Considering that the model's artifacts represent evaluating collaborative environments, we can conclude that the model has content validity.

After this refinement, we started planning and executing three case studies (WOHLIN et al., 2012; YIN, 2009) through a large-scale evaluation of our assessment model to evaluate the instrument reliability and construct validity (TROCHIM; DONNELLY, 2001; DEVELLIS, 2016).

The first case study scenario, described in Chapter 8, was designed to refine the model and adjust/adequate the artifacts. In this scenario, we evaluated the collaboration aspects provided by a single collaborative environment, considering a small sample of participants. As a result, suitable indicators from the perspective of demographic data and IRT parameterization were found.

In the second case study scenario, described in Chapter 9, we evaluated a set of collaborative applications to enable the extraction of results for each of the evaluated environments and thus verify the behavior of the evaluation model in each. As a result, the awareness quality scale was established considering the participants' ability to identify awareness information; consequently, higher scores indicate that the evaluated environments easily support awareness mechanisms, while participants with higher ability scores can identify existing awareness mechanisms properly.

In the third scenario, described in Chapter 10, we expose the model's artifacts to HCI and collaborative system examiners' appreciation to verify the adequacy of the process, its activities and related artifacts to evaluate the awareness support in collaborative environments from the examiners' perspective. As a result, the model can be replicated fully or partially in other scenarios and contexts by selecting the dimensions, categories, and awareness mechanisms relevant to the scenario and adapting the awareness support scale.

Finally, we compiled all observations obtained through all evaluation scenarios to generate the Global Awareness Scale (described in Chapter 11). We assumed the estimation of participants' abilities and items' parameters as a calibration strategy and the IRT multi-group estimation method (CHALMERS, 2012). In this approach, calibration was performed by maximum likelihood analysis for polytomous data (Samejima's gradual scale (SAMEJIMA, 1969)) using the Metropolis-Hastings Robbins-Monro (MHRM) algorithm approach (CAI, 2010). As a result, the global scale validation was very positive, considering the internal consistency, reliability, and dimensionality (EFA e CFA) perspectives. In summary, exploratory factor analysis (EFA) indicated a strong tendency towards the one-dimensional model (latent root criterion) (HAIR et al., 2009), legitimating the correlation between the assessment items and the observed latent trail; the confirmatory factor analysis (CFA) results demonstrate the instrument construct validity (FORNELL; LARCKER, 1981; HAIR et al., 2009; TAVAKOL; WETZEL, 2020).

12.2 LIMITATIONS

This research assumes the awareness taxonomy's **transferability** and assessment model's **generalizability**. Transferability means that the knowledge generated by research is not generalized; it is only transferred to a similar context (transferring knowledge generated from a sending context to a receiving context) (POLIT; BECK, 2010). Thus, the reader must identify how much the knowledge applies to another problem.

Generalizability is the degree to which the results can be applied to a broader context. Research results are considered generalizable when the findings can be applied to most contexts, most people, most of the time (POLIT; BECK, 2010).

In the taxonomy, the awareness mechanisms described were judged applicable to general-purpose collaborative environments; however, they may not be a complete set of support resources in a more specific collaboration context (i.e., disruptive, ubiquitous, or immersive environment). Therefore, awareness mechanisms or strategies can consider other contextual information depending on the application domain.

In the assessment model validation, we exposed the model from the perspective of three distinct samples (population): collaborative systems and HCI experts (by expert panel scenario), examinees, and examiners (by three case study scenarios). In this case, the results were generalized considering this population, and the results can be generalizable since the sample obtained is representative to other scenarios.

Finally, the awareness assessment model was developed based on IRT, resulting in an adaptive/flexible assessment approach; thus, new assessment items (i.e., awareness mechanisms or other HCI aspects) can be included in the model. It is imperative to notice that the model must be calibrated for the desired items whenever new items are incorporated. Furthermore, the IRT model calibration stage can be relatively complex if the examiner has insufficient statistical knowledge. General statistical analysis tools, such as R, are necessary to execute the model calibration and analysis scripts.

12.3 FUTURE WORK

Resulting from this work, we encourage research in the following direction:

- The development of an assessment tool or system allows a dashboard view of the analysis carried out in this work. This allows a person to obtain a more detailed description of each mechanism or dimension of awareness support assessed.

- The development of model evaluations in other scenarios or contexts of use, such as users with different skill levels, limitations, or special needs. Furthermore, evaluating how other examiners use the model while evaluating collaborative systems may be interesting.

- Models for evaluating support awareness in computer-supported collaborative learning (CSCL) are promising; likewise, research supporting context-aware or other specific dimensions of awareness will be fruitful.

- Expand the analysis by discussing other awareness or software quality aspects in the same assessment, for example, incorporating usability issues or other HCI assessment mechanisms into a more general approach.

- Establish an awareness model towards supporting the design and development of collaborative applications, for example, by a framework, method, or specification of awareness support mechanisms. Awareness design remains far from being solved, and the gap between theory and design remains open; there is a lack of established guidelines, specific tools, and systematic support for the modeling and designing awareness requirements.

- Converge evaluation solutions focused on the emerging new collaborative settings (i.e., disruptive devices, ubiquitous computing, immersive environments, virtual worlds) where the implicit and multimodal interaction styles, including the use of haptic devices, affective user interfaces, and other new devices at the cutting edge of technology, are widely used instead of more classical user interfaces and collaboration metaphors.

- Extend the specification of this awareness assessment model in compliance with other assessment/process quality standards, like ISO/IEC 33000 family, allowing its adoption together with other processes or software product assessment standards.

- Direct efforts to establish guidelines on prioritizing awareness assessment items to help examiners choose which items are most likely to be evaluated, depending on each concrete scenario (i.e., adapting the assessment items or adding new ones, considering non-functional aspects or use restrictions, or considering other quality attributes).

- The development of a reference guide encompassing the evaluation process and conceptual vision to facilitate its adoption by novices or examiners with little knowledge about collaboration, awareness, or collaborative environments assessment aspects.

- The development of a computational assessment tool using the Computer-based Adaptive Test (CAT) strategies (WEISS, 1982; STRAETMANS; EGGEN, 1998), reducing, for example, the number of assessment items or time required.

- Evaluating the awareness from a qualitative perspective can be promising. Furthermore, we encourage future investigations related to the ‘how much” awareness information (KIRSCH-PINHEIRO; DE LIMA; BORGES, 2003) should be provided in the collaborative interface.

- Extend the specification of this awareness assessment model to consider the persona, historical, and boundary awareness facets, as described in the awareness taxonomy 4.

12.4 RELATED PUBLICATIONS

This research results in the following publications.

- MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. Awareness Support in Collaborative System: Reviewing last 10 years of CSCW Research. In: IEEE. 2022 IEEE 25th International Conference on Computer-Supported Cooperative Work in Design (CSCWD). IEEE, 2022. P. 564–569. DOI: 10.1109/cscwd54268.2022.9776091.
- MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. Towards an Awareness Taxonomy. In: IEEE. 2022 IEEE 25th International Conference on Computer-Supported Cooperative Work in Design (CSCWD). IEEE, 2022. P. 495–500. DOI: 10.1109/cscwd54268.2022.9776129.
- MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. The Awareness Assessment Model: measuring the awareness and collaboration support over the participant's perspective. In: Anais do XVIII Simpósio Brasileiro de Sistemas Colaborativos, SBSC. SBC, 2023. P. 30-43.
- MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. The Awareness Assessment Model: Measuring Awareness and Collaboration Support Over Participant's Perspective. In: Universal Access in the Information Society (UAIS). DOI: 10.1007/s10209-024-01110-5.
- MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. The Awareness Assessment Model repository. Version 1.0. Zenodo, Aug. 2023. DOI: 10.5281/zenodo.8298950.

REFERENCES

AGOGO, David; HESS, Traci J. “How does tech make you feel?” a review and examination of negative affective responses to technology use. **European Journal of Information Systems**, Taylor & Francis, v. 27, n. 5, p. 570–599, 2018. DOI: 10.1080/0960085X.2018.1435230. Cit. on pp. 51, 138.

ALTENBURGER, T. et al. Groupware requirements modelling for adaptive user interface design. English. **eWork and eBusiness in Architecture, Engineering and Construction - Proceedings of the European Conference on Product and Process Modelling 2012, ECPPM 2012**, Reykjavik, Iceland, v. 1, n. 1, p. 825–832, 2012. DOI: 10.1201/b12516-131. Cit. on pp. 82, 95, 272–275, 277, 278, 280.

ANDRADE, Dalton Francisco de; TAVARES, Heliton Ribeiro; VALLE, Raquel da Cunha. Teoria da Resposta ao Item: conceitos e aplicações. **ABE, Sao Paulo**, 2000. Cit. on pp. 37, 53, 54, 56, 61, 63–65, 67, 71, 76, 131, 137, 176, 184, 192, 209, 225.

ANDRICH, David. A rating formulation for ordered response categories. **Psychometrika**, Springer, v. 43, n. 4, p. 561–573, Dec. 1978. ISSN 1860-0980. DOI: 10.1007/bf02293814. Cit. on p. 56.

ANTUNES, Pedro et al. Reviewing the quality of awareness support in collaborative applications. English. **Journal of Systems and Software**, elsevier, v. 89, n. 1, p. 146–169, 2014. ISSN 0164-1212. DOI: 10.1016/j.jss.2013.11.1078. Cit. on pp. 33, 34, 49, 73–75, 82, 84, 90, 102, 114, 272–280, 285–287.

ANTUNES, Pedro et al. Structuring dimensions for collaborative systems evaluation. English. **ACM Computing Surveys**, v. 44, n. 2, p. 1–28, 2012. ISSN 0360-0300. Cit. on pp. 82, 95, 272, 273, 275, 276.

AYALA, Rafael J. The IRT tradition and its applications. **The Oxford handbook of quantitative methods**, Oxford University Press New York, NY, v. 1, 144e169, 2013. DOI: 10.1093/oxfordhb/9780199934874.013.0008. Cit. on pp. 54, 176.

AYALA, Rafael J. **The theory and practice of item response theory**. New York, NY, USA: Guilford Publications, 2013. Cit. on pp. 53, 54.

BAILEY, Kenneth D. **Typologies and taxonomies: An introduction to classification techniques**. London, UK: Sage Publications, 1994. v. 102. Cit. on pp. 42, 50, 102, 242.

BAKER, Frank B. **The basics of item response theory**. University of Wisconsin, WI, USA: ERIC, 2001. Cit. on pp. 53, 55, 56, 58, 60, 63, 186, 227.

BAKER, Frank B; KIM, Seock-Ho. **Item response theory: Parameter estimation techniques**. Madison, WI, USA: CRC press, 2004. Cit. on pp. 61, 128, 129.

BAKER, Frank B; KIM, Seock-Ho. **The basics of item response theory using R**. Madision, WI, USA: Springer, 2017. DOI: 10.1007/978-3-319-54205-8. Cit. on pp. 43, 53, 56, 58, 129–131, 186, 227.

BARBOSA, Simone; SILVA, Bruno. **Interação humano-computador**. Rio de Janeiro, RJ: Elsevier Brasil, 2010. Cit. on p. 117.

BARTOLUCCI, Francesco. A class of multidimensional IRT models for testing unidimensionality and clustering items. **Psychometrika**, v. 72, n. 2, p. 141–157, June 2007. ISSN 1860-0980. DOI: 10.1007/s11336-005-1376-9. Cit. on p. 38.

BASILI, VR. Software modeling and measurement: The Goal/Question/Metric paradigm. **Technical Report**, Department of Computer Science, University of Maryland, 1992. Cit. on pp. 43, 117, 122, 139.

BEECHAM, Sarah et al. Using an expert panel to validate a requirements process improvement model. **Journal of Systems and Software**, Elsevier, v. 76, n. 3, p. 251–275, 2005. DOI: 10.1016/j.jss.2004.06.004. Cit. on pp. 12, 13, 43, 135, 138, 243.

BELKADI, Farouk et al. A situation model to support awareness in collaborative design. English. **International Journal of Human Computer Studies**, v. 71, n. 1, p. 110–129, 2013. ISSN 1071-5819. DOI: 10.1016/j.ijhcs.2012.03.002. Cit. on pp. 82, 91, 95, 274, 276–278, 280.

BERKMAN, Mehmet Iker; KARAHOCA, Dilek; KARAHOCA, Adem. A Measurement and Structural Model for Usability Evaluation of Shared Workspace Groupware. English. **International Journal of Human-Computer Interaction**, v. 34, n. 1, p. 35–56, 2018. ISSN 1044-7318. Cit. on pp. 73, 82, 90, 92.

BICHLER, Martin. Design science in information systems research. **Wirtschaftsinformatik**, Springer, v. 48, n. 2, p. 133–135, 2006. Cit. on pp. 10, 40.

BORGES, José Wicto Pereira et al. Measuring the quality of life in hypertension according to Item Response Theory. **Revista de Saúde Pública**, FapUNIFESP (SciELO), v. 51, n. 0, 2017. ISSN 0034-8910. DOI: 10.1590/s1518-8787.2017051006845. Cit. on p. 53.

BORTOLOTTI, Silvana Ligia Vincenzi et al. Relevance and advantages of using the item response theory. **Quality & Quantity**, Springer Science and Business Media LLC, v. 47, n. 4, p. 2341–2360, Mar. 2012. ISSN 1573-7845. DOI: 10.1007/s11135-012-9684-5. Cit. on pp. 54, 76.

BRAVO, Crescencio et al. Modeling Awareness Requirements in Groupware: From Cards to Diagrams. **IEEE Transactions on Human-Machine Systems**, Institute of Electrical and Electronics Engineers (IEEE), p. 1–12, 2023. ISSN 2168-2305. DOI: 10.1109/thms.2023.3332592. Cit. on pp. 9, 36, 49, 50, 95.

BREZILLON, P. et al. Context-based awareness in Group Work. In: PROCEEDING of the 17th International FLAIRS Conference. Miami, FL, USA: AAAI Press, 2004. Cit. on pp. 34, 48.

BROWN, Timothy A. **Confirmatory factor analysis for applied research**. New York, NY, USA: Guilford publications, 2015. Cit. on pp. 72, 192, 232.

BULLEN, Christine V; JOHANSEN, Robert. Groupware, a key to managing business teams? **Cambridge, Mass.: Center for Information Systems Research, Sloan School of**, 1988. Cit. on p. 88.

CAI, Li. Metropolis-Hastings Robbins-Monro Algorithm for Confirmatory Item Factor Analysis. **Journal of Educational and Behavioral Statistics**, American Educational Research Association (AERA), v. 35, n. 3, p. 307–335, June 2010. ISSN 1935-1054. DOI: 10.3102/1076998609353115. Cit. on pp. 14, 209, 244.

CHADHA, Narender Kumar. **Applied psychometry**. Pennsylvania State University, State College, PA, USA: SAGE Publications India, 2009. DOI: 10.4135/9788132108221. Cit. on pp. 51–53.

CHALMERS, R. Philip. mirt: A Multidimensional Item Response Theory Package for theREnvironment. **Journal of Statistical Software**, Foundation for Open Access Statistic, v. 48, n. 6, 2012. ISSN 1548-7660. DOI: 10.18637/jss.v048.i06. Cit. on pp. 14, 130, 137, 164, 191, 209, 232, 244.

COLLAZOS, Cesar A. et al. Descriptive theory of awareness for groupware development. English. **Journal of Ambient Intelligence and Humanized Computing**, v. 10, n. 12, p. 4789–4818, 2019. ISSN 1868-5137. Cit. on pp. 9, 36, 39, 47–49, 73–75, 82, 90, 91, 93, 95, 102, 114.

COLVIN, Kimberly F; GORGUN, Guher. Collapsing Scale Categories: Comparing the Psychometric Properties of Resulting Scales. University of Massachusetts Amherst, 2020. DOI: 10.7275/17315702. Available from: <https://openpublishing.library.umass.edu/pare/article/id/1608/>. Cit. on p. 146.

CRONBACH, Lee J. Coefficient alpha and the internal structure of tests. **psychometrika**, Springer, v. 16, n. 3, p. 297–334, 1951. Cit. on pp. 12, 44, 70, 136.

CRUZ, Armando et al. Towards an overarching classification model of CSCW and groupware: A socio-technical perspective. English. In: LECTURE Notes in Computer Science. Raesfeld, Germany: Elsevier, 2012. P. 41–56. DOI: 10.1007/978-3-642-33284-5_4. Cit. on pp. 35, 82, 91–95, 274, 275, 277.

DE ARAÚJO, Luciana Pereira et al. Heuristic evaluation for mobile groupware: Evaluating two audience response systems. **Journal of Applied Computing Research**, v. 3, n. 2, p. 64–77, 2014. Cit. on pp. 73, 82, 90–92, 94, 102.

DECOUCHANT, Dominique et al. Adapting groupware systems to changes in the collaborator's context of use. English. **Expert Systems with Applications**, v. 40, n. 11, p. 4446–4462, 2013. ISSN 0957-4174. DOI: 10.1016/j.eswa.2013.01.043. Cit. on p. 82.

DEMARS, Christine. **Item response theory**. New York, NY, USA: Oxford University Press, 2010. DOI: 10.1093/acprof:oso/9780195377033.001.0001. Cit. on pp. 44, 55, 56, 59, 60, 63, 70, 136.

DEVELLIS, Robert F. **Scale development: Theory and applications**. 4th. London, UK: Sage Publications, Inc, 2016. v. 26. (Applied Social Research Methods 26). ISBN 9781506341569. Cit. on pp. 13, 43, 44, 70, 122, 131, 136, 159, 164, 191, 201, 210, 231, 232, 243.

DORANS, Neil J; MOSES, Tim P; EIGNOR, Daniel R. Principles and practices of test score equating. **ETS Research Report Series**, Wiley Online Library, v. 2010, n. 2, p. 1–41, 2010. Cit. on p. 65.

DOURISH, Paul. **Where the action is: the foundations of embodied interaction**. Cambridge, MA, USA: MIT Press, 2004. ISBN 0262041960. DOI: 10.7551/mitpress/7221.001.0001. Cit. on pp. 34, 48.

DOURISH, Paul; BELLOTTI, Victoria. Awareness and Coordination in Shared Workspaces. In: PROCEEDINGS of the 1992 ACM Conference on Computer-Supported Cooperative Work. Toronto, Ontario, Canada: Association for Computing Machinery, 1992. (CSCW '92), p. 107–114. DOI: 10.1145/143457.143468. Cit. on pp. 8, 33, 49.

DYBA, Tore. An instrument for measuring the key factors of success in software process improvement. **Empirical software engineering**, Springer, v. 5, p. 357–390, 2000. Cit. on p. 138.

DYBÅ, Tore; DINGSØYR, Torgeir. Empirical studies of agile software development: A systematic review. English. **Information and Software Technology**, v. 50, n. 9-10, p. 833–859, 2008. ISSN 0950-5849. DOI: 10.1016/j.infsof.2008.01.006. Cit. on p. 96.

DYBÅ, Tore; DINGSØYR, Torgeir. Strength of Evidence in Systematic Reviews in Software Engineering. In: PROCEEDINGS of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement. Kaiserslautern, Germany: Association for Computing Machinery, 2008. (ESEM '08), p. 178–187. DOI: 10.1145/1414004.1414034. Cit. on p. 96.

EL EMAM, K.; BIRK, A. Validating the ISO/IEC 15504 measure of software requirements analysis process capability. **IEEE Transactions on Software Engineering**, Institute of Electrical and Electronics Engineers (IEEE), v. 26, n. 6, p. 541–566, June 2000. ISSN 2326-3881. DOI: 10.1109/32.852742. Cit. on p. 138.

ELLIS, Clarence; GIBBS, Simon; REIN, Gail. Groupware some issues and experiences. English. **Communications of the ACM**, ACM New York, NY, USA, v. 34, n. 1, p. 38–58, 1991. ISSN 0001-0782. DOI: 10.1145/99977.99987. Cit. on pp. 8, 33, 46, 47, 88.

EMBRETSON, Susan E. A multidimensional latent trait model for measuring learning and change. **Psychometrika**, Springer, v. 56, p. 495–515, 1991. DOI: 10.1007/bf02294487. Cit. on p. 52.

ESPIRITO SANTO, Diogo Alberto et al. Quality assessment of awareness support in agile collaborative tools. English. In: PROCEEDINGS - 2018 44th Latin American Computing Conference, CLEI 2018. São Paulo, SP, BR: Elsevier, 2018. P. 21–30. DOI: 10.1109/clei.2018.00013. Cit. on pp. 73, 74, 82, 84, 102, 114, 272–277, 279, 280, 285–287.

FAUZI, S.S.M.; SOBRI, W.A.W.M.; SUALI, A.J. Synthesising the literature on the benefits and impact of group awareness in group work. **Pertanika Journal of Social Sciences and Humanities**, Universiti Putra Malaysia, v. 25, February, p. 119–130, 2017. ISSN 0128-7702. Cit. on p. 82.

FAYERS, Peter. Item Response Theory for Psychologists. **Quality of Life Research**, Springer Science and Business Media LLC, v. 13, n. 3, p. 715–716, Apr. 2004. ISSN 0962-9343. DOI: 10.1023/b:qure.0000021503.45367.f2. Cit. on p. 54.

FELDT, Leonard S; CHARTER, Richard A. Estimating the reliability of a test split into two parts of equal or unequal length. **Psychological methods**, American Psychological Association, v. 8, n. 1, p. 102, 2003. DOI: 10.1037/1082-989x.8.1.102. Cit. on p. 63.

FORERO, Carlos G.; MAYDEU-OLIVARES, Alberto. Estimation of IRT graded response models: Limited versus full information methods. **Psychological Methods**, American Psychological Association (APA), v. 14, n. 3, p. 275–299, 2009. DOI: 10.1037/a0015825. Cit. on p. 57.

FORNELL, Claes; LARCKER, David F. Evaluating structural equation models with unobservable variables and measurement error. **Journal of marketing research**, Sage Publications Sage CA: Los Angeles, CA, v. 18, n. 1, p. 39–50, 1981. Cit. on pp. 15, 196, 201, 236, 241, 244.

FUKS, Hugo; GEROSA, Marco; RAPOSO, Alberto. Engineering Groupware for E-Business. In: FIRST Seminar on Advanced Research in Electronic Business (EBR2002). Rio de Janeiro, RJ, BR: Citeseer, 2002. P. 78–84. Cit. on p. 47.

FUKS, Hugo; RAPOSO, Alberto; GEROSA, Marco; PIMENTAL, Mariano, et al. The 3C collaboration model. In: ENCYCLOPEDIA of E-collaboration. Pennsylvania, USA: IGI Global, 2008. DOI: 10.4018/978-1-59904-000-4.ch097. Cit. on pp. 47, 48.

FUKS, Hugo; RAPOSO, Alberto; GEROSA, Marco Aurélio; PIMENTEL, Mariano, et al. Inter-and intra-relationships between communication coordination and cooperation in the scope of the 3C Collaboration Model. In: 2008 12th International Conference on Computer Supported Cooperative Work in Design. Xi'an, China: IEEE, 2008. P. 148–153. DOI: 10.1109/cscwd.2008.4536971. Cit. on p. 47.

FUKS, Hugo; RAPOSO, Alberto B; GEROSA, Marco A; LUCENA, Carlos JP. Applying the 3C model to groupware development. **International Journal of Cooperative Information Systems**, World Scientific, v. 14, 02n03, p. 299–328, 2005. DOI: 10.1142/s0218843005001171. Cit. on pp. 47, 48.

GALLARDO, Jesus; BRAVO, Crescencio; MOLINA, Ana Isabel. A framework for the descriptive specification of awareness support in multimodal user interfaces for collaborative activities. English. **Journal on Multimodal User Interfaces**, v. 12, n. 2, p. 145–159, 2018. ISSN 1783-7677. Cit. on pp. 9, 33, 36, 49, 82, 114, 274, 276, 277, 285–287.

GALLARDO, Jesús; MOLINA, Ana Isabel, et al. An ontological conceptualization approach for awareness in domain-independent collaborative modeling systems: Application to a model-driven development method. In: 2, p. 1099–1118. DOI: 10.1016/j.eswa.2010.05.005. Cit. on pp. 73, 74, 82, 90, 102, 272–275, 277.

GEORGE, Joey F. Groupware. In: BIDGOLI, Hossein (Ed.). **Encyclopedia of Information Systems**. New York, NY, USA: Elsevier, 2003. P. 509–518. ISBN 978-0-12-227240-0. DOI: 10.1016/B0-12-227240-4/00084-8. Cit. on pp. 8, 33, 46.

GEROSA, M.A.; FUKS, H.; LUCENA, C.P.J. Suporte a percepção em Ambientes Digitais de Aprendizagem. **Revista Brasileira de Informá1tica na Educação, Sociedade Brasileira de Computação**, v. 11, p. 75–85, 2003. Cit. on p. 49.

GIL, Antonio Carlos et al. **Como elaborar projetos de pesquisa**. São Paulo, SP, BR: Atlas São Paulo, 2002. v. 4. Cit. on p. 52.

GREENBERG, Saul. Collaborative interfaces for the web. **Human factors and web development**, Citeseer, v. 18, p. 241–254, 1997. Cit. on pp. 34, 35, 49.

GREIF, Irene (Ed.). **Computer-Supported Cooperative Work: A Book of Readings**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988. ISBN 0934613575. Cit. on pp. 46, 47.

GROSS, Tom. Supporting effortless coordination: 25 years of awareness research. English. **Computer Supported Cooperative Work**, v. 22, n. 4-6, p. 425–474, 2013.

ISSN 0925-9724. DOI: 10.1007/s10606-013-9190-x. Cit. on pp. 8, 9, 33, 34, 39, 46, 49, 82, 92, 99.

GUTWIN, Carl; BARJAWI, Mutasem; PINELLE, David. 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. Sanibel Island, Florida, USA: Association for Computing Machinery, 2016. (GROUP '16), p. 277–286. DOI: 10.1145/2957276.2957315. Cit. on pp. 47, 83.

GUTWIN, Carl; BATEMAN, Scott, 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. Portland, Oregon, USA: Association for Computing Machinery, 2017. (CSCW '17), p. 2221–2235. DOI: 10.1145/2998181.2998226. Cit. on pp. 73, 82, 90, 92, 280.

GUTWIN, Carl; GREENBERG, Saul. A descriptive framework of workspace awareness for real-time groupware. English. **Computer Supported Cooperative Work**, v. 11, n. 3-4, p. 411–446, 2002. ISSN 0925-9724. DOI: 10.1023/a:1021271517844. Cit. on pp. 34, 35, 84, 99, 105, 242.

GUTWIN, Carl; GREENBERG, Saul; ROSEMAN, Mark. Workspace Awareness in Real-Time Distributed Groupware: Framework, Widgets, and Evaluation. In: SASSE, Martina Angela; CUNNINGHAM, R. Jim; WINDER, Russel L. (Eds.). **People and Computers XI**. Springer-Verlag: Springer London, 1996. (HCI'96), p. 281–298. DOI: 10.1007/978-1-4471-3588-3_18. Cit. on pp. 34, 88, 89.

HADDAWAY, Neal R. et al. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. **Campbell Systematic Reviews**, John Wiley & Sons, Ltd, v. 18, n. 2, e1230, June 2022. ISSN 1891-1803. DOI: 10.1002/cl2.1230. Cit. on p. 79.

HAIR, Joseph F et al. **Análise multivariada de dados**. Porto Alegre, RS, BR: Bookman editora, 2009. Cit. on pp. 12, 14, 15, 44, 71–73, 136, 145, 169, 192, 193, 195, 196, 201, 232, 233, 236, 239, 241, 244.

HAKIM, Catherine. **Research design: Strategies and choices in the design of social research**. London, UK: Allen and Unwin, 1987. Cit. on p. 138.

HARVEY, Robert J; HAMMER, Allen L. Item response theory. **The Counseling Psychologist**, Sage Publications Sage CA: Thousand Oaks, CA, v. 27, n. 3, p. 353–383, 1999. Cit. on p. 54.

HERSKOVIC, Valeria; OCHOA, Sergio F; PINO, José A; NEYEM, H Andrés. The Iceberg Effect: Behind the User Interface of Mobile Collaborative Systems. **JUCS - Journal of Universal Computer Science**, v. 17, n. 2, p. 183–201, 2011. Cit. on pp. 73, 82, 90, 93, 94, 272.

HERSKOVIC, Valeria; OCHOA, Sergio F.; PINO, Jose A.; ANTUNES, Pedro, et al. Identifying the awareness mechanisms for mobile collaborative applications. English. In: LECTURE Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Wellington, New Zealand: Springer, 2013. P. 241–256. DOI: 10.1007/978-3-642-41347-6_18. Cit. on pp. 82, 272–275, 277.

HEVNER, Alan R. A three cycle view of design science research. **Scandinavian journal of information systems**, v. 19, n. 2, p. 4, 2007. Cit. on p. 40.

HINCAPIE-RAMOS, Juan David; VOIDA, Stephen; MARK, Gloria. A design space analysis of availability-sharing systems. English. In: UIST'11 - Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. Santa Barbara, CA, United states: Elsevier, 2011. P. 85–95. DOI: 10.1145/2047196.2047207. Cit. on pp. 82, 94.

HINKELMANN, Klauss; KEMPTHORNE, Oscar. **Design and analysis of experiments: Advanced experimental design**. Hoboken, NJ, USA: Wiley-Interscience, 2005. v. 2. DOI: 10.1002/0471709948. Cit. on pp. 127, 144, 169, 240.

HOLLAND, Paul W.; DORANS, Neil J.; PETERSEN, Nancy S. 6 Equating Test Scores. In: RAO, C.R.; SINHARAY, S. (Eds.). **Psychometrics**. Princeton, NJ, USA: Elsevier, 2006. v. 26. (Handbook of Statistics). P. 169–203. DOI: 10.1016/S0169-7161(06)26006-1. Cit. on pp. 40, 66, 67, 209.

HUMMEL, Dennis; SCHACHT, Silvia; MAEDCHE, Alexander. Determinants of Multi-Channel Behavior: Exploring Avenues for Future Research in the Services Industry. In: ICIS. Dublin, Ireland: Fraunhofer-Gesellschaft, 2016. Cit. on pp. 51, 138.

IDRUS, Zainura et al. Social awareness: The power of digital elements in collaborative environment. English. **WSEAS Transactions on Computers**, World Scientific, Engineering Academy, and Society (WSEAS) Stevens Point, v. 9, n. 6, p. 644–653, 2010. ISSN 1109-2750. Cit. on pp. 82, 274, 275, 277.

IIVARI, Juhani; VENABLE, John. Action research and design science research - Seemingly similar but decisively dissimilar. In: NEWELL, Susan et al. (Eds.). **17th European Conference on Information Systems, ECIS**. Verona, Italy: Association for Information Systems, 2009. P. 1642–1653. Cit. on p. 40.

STANDARDIZATION, International Organization for. **Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Evaluation process**. v. 1. Geneva, CH, Mar. 2011. Cit. on pp. 117, 243.

IZQUIERDO, Isabel; OLEA, Julio; ABAD, Francisco José. Exploratory factor analysis in validation studies: uses and recommendations. en. **Psicothema**, Spain, v. 26, n. 3, p. 395–400, 2014. Cit. on pp. 12, 44, 72, 136, 192, 193, 232.

JONASSON, Niclas. Designing for Collaboration: Interface Guidelines for Improving Co-Located Collaboration on Tabletop Displays. English. In: PROCEEDINGS of Umea's 14th Student Conference in Computing Science - USCCS 2010. Umea, Sweeden: Umea University, 2010. P. 61–71. Cit. on pp. 82, 93.

KHALILIJAFARABAD, Ahmad; HELFERT, Markus; GE, Mouzhi. Developing a Data Quality Research Taxonomy-an organizational perspective. In: ICIQ: International Conference on Information Quality. Ciudad Real, Spain: Universidad de Castilla-La Mancha, 2016. P. 176–186. Cit. on pp. 51, 138.

KIM, Kyung Tae et al. Hugin: A framework for awareness and coordination in mixed-presence collaborative information visualization. English. In: ACM International Conference on Interactive Tabletops and Surfaces, ITS 2010. New York, NY, USA: ACM, 2010. P. 231–240. Cit. on pp. 73, 82, 90, 93, 94, 272, 280.

KIRSCH-PINHEIRO, Manuele; DE LIMA, José Valdeni; BORGES, Marcos RS. A framework for awareness support in groupware systems. **Computers in Industry**, Elsevier, v. 52, n. 1, p. 47–57, 2003. Cit. on pp. 34, 246.

KITCHENHAM, B.A.; PFLEEGER, S.L., et al. Preliminary guidelines for empirical research in software engineering. **IEEE Transactions on Software Engineering**, Institute of Electrical and Electronics Engineers (IEEE), v. 28, n. 8, p. 721–734, Aug. 2002. ISSN 0098-5589. DOI: 10.1109/tse.2002.1027796. Cit. on p. 139.

KITCHENHAM, Barbara; BRERETON, Pearl. A systematic review of systematic review process research in software engineering. **Information and Software Technology**, v. 55, n. 12, p. 2049–2075, 2013. ISSN 0950-5849. DOI: 10.1016/j.infsof.2013.07.010. Cit. on p. 96.

KITCHENHAM, Barbara; CHARTERS, Stuart. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. University of Durham, Durham, UK, July 2007. Cit. on pp. 10, 42, 77, 96, 98, 242.

KOLEN, Michael J.; BRENNAN, Robert L. **Test Equating, Scaling, and Linking**. New York, NY, USA: Springer New York, 2014. DOI: 10.1007/978-1-4939-0317-7. Cit. on pp. 68, 69.

LAUESEN, Soren; VINTER, Otto. Preventing Requirement Defects: An Experiment in Process Improvement. **Requirements Engineering**, Springer Science and Business Media LLC, v. 6, n. 1, p. 37–50, Feb. 2001. ISSN 0947-3602. DOI: 10.1007/pl00010355. Cit. on pp. 138, 139.

LEDESMA, Rubén Daniel; VALERO-MORA, Pedro. Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. **Practical assessment, research, and evaluation**, v. 12, n. 1, p. 2, 2019. Cit. on pp. 72, 192, 233.

LINACRE, J M. Investigating rating scale category utility. en. **J Outcome Meas**, United States, v. 3, n. 2, p. 103–122, 1999. Cit. on pp. 146, 165, 202.

LOPEZ, Gustavo; GUERRERO, Luis 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. Portland, Oregon, USA: Association for Computing Machinery, 2017. (CSCW '17), p. 808–820. DOI: 10.1145/2998181.2998281. Cit. on pp. 9, 33, 36, 49, 74, 98.

LORD, FM. **Applications of item response theory to practical testing problems**. New York, NY, USA: Erlbaum, 1980. Cit. on pp. 68, 69.

MAKRANSKY, Guido; LILLEHOLT, Lau; AABY, Anders. Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. **Computers in Human Behavior**, v. 72, p. 276–285, 2017. ISSN 0747-5632. DOI: 10.1016/j.chb.2017.02.066. Cit. on p. 53.

MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. 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). Hangzhou, China: IEEE, 2022. P. 564–569. DOI: 10.1109/cscwd54268.2022.9776091. Cit. on p. 73.

MANTAU, Márcio José; BENITTI, Fabiane Barreto Vavassori. **The Awareness Assessment Model repository**. Version 1.0. Personal Collection: Zenodo, Sept. 2023. DOI: 10.5281/zenodo.8298950. Cit. on pp. 130, 143, 145, 169, 202, 204, 210, 243.

MANTAU, Márcio José; BERKENBROCK, Carla Diacui Medeiros; BERKENBROCK, Gian Ricardo. Análise de Requisitos de Percepção em um Groupware móvel síncrono. Portuguese. **Revista de Sistemas de Informação da FSMA**, v. 1, n. 19, p. 16–33, 2017. ISSN 1044-7318. Cit. on pp. 33–35, 49, 73, 82, 90–94, 102, 279, 280.

MANTAU, Márcio José; BERKENBROCK, Carla Diacui Medeiros; BERKENBROCK, Gian Ricardo. 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). National Tsing Hua University, Hsinchu, Taiwan: IEEE, 2014. (CSCWD'14), p. 563–568. DOI: 10.1109/cscwd.2014.6846906. Cit. on pp. 35, 73, 82, 90, 91, 93–95, 167, 278–280.

MASTERS, Geoff N. A rasch model for partial credit scoring. **Psychometrika**, v. 47, n. 2, p. 149–174, June 1982. ISSN 1860-0980. DOI: 10.1007/BF02296272. Cit. on p. 57.

MATOOK, Sabine; INDULSKA, Marta. Improving the quality of process reference models: A quality function deployment-based approach. **Decision support systems**, v. 47, n. 1, p. 60–71, 2009. DOI: 10.1016/j.dss.2008.12.006. Cit. on p. 114.

MOLINA, Ana Isabel et al. Assessing the awareness mechanisms of a collaborative programming support system. **Dyna**, v. 82, n. 193, p. 212–222, 2015. DOI: 10.15446/dyna.v82n193.53497. Cit. on pp. 48, 74, 75.

NEYEM, Andres et al. A reusable structural design for mobile collaborative applications. English. **Journal of Systems and Software**, v. 85, n. 3, p. 511–524, 2012. ISSN 0164-1212. DOI: 10.1016/j.jss.2011.05.046. Cit. on pp. 82, 93, 94.

NGUYEN, Tam H. et al. An Introduction to Item Response Theory for Patient-Reported Outcome Measurement. **The Patient - Patient-Centered Outcomes Research**, Springer Science and Business Media LLC, v. 7, n. 1, p. 23–35, Jan. 2014. ISSN 1178-1661. DOI: 10.1007/s40271-013-0041-0. Cit. on p. 39.

NICKERSON, Robert C; VARSHNEY, Upkar; MUNTERMANN, Jan. A method for taxonomy development and its application in information systems. **European Journal of Information Systems**, Springer, v. 22, n. 3, p. 336–359, 2013. DOI: 10.1057/ejis.2012.26. Cit. on pp. 11, 42, 50, 51, 102, 104, 113, 138, 142, 203, 243.

NICOLACI-DA-COSTA, A.M.; PIMENTEL, M. Sistemas colaborativos para uma nova sociedade e um novo ser humano. In: PIMENTEL, M.; FUKS, H. (Eds.). **Sistemas Colaborativos**. São Paulo, SP, BR: Elsevier, 2011. P. 3–15. Cit. on p. 47.

NIEMANTSVERDRIET, Karin et al. Designing for Awareness in Interactions with Shared Systems: The DASS Framework. **ACM Transaction on Computer-Human Interaction**, Association for Computing Machinery, New York, NY, USA, v. 26, n. 6, p. 1–41, Nov. 2019. ISSN 1073-0516. DOI: 10.1145/3338845. Cit. on pp. 36, 48, 73, 75, 82, 90–92, 94, 95, 102, 114, 274–277, 285–287.

NUNNALLY, Jum C; BERNSTEIN, Ira. **Psychometric Theory**. 3. ed. Maidenhead, UK: McGraw Hill Higher Education, Nov. 1993. (McGraw-Hill Series in Psychology). Cit. on pp. 69, 70, 135.

OLINO, Thomas M. et al. Measuring depression using item response theory: an examination of three measures of depressive symptomatology. **International Journal of Methods in Psychiatric Research**, Wiley, v. 21, n. 1, p. 76–85, Jan. 2012. ISSN 1557-0657. DOI: 10.1002/mpr.1348. Cit. on p. 53.

OLSON, Gary M.; OLSON, Judith S. Chapter 59 - Research on Computer Supported Cooperative Work. In: HELANDER, Marting G.; LANDAUER, Thomas K.; PRABHU, Prasad V. (Eds.). **Handbook of Human-Computer Interaction (Second Edition)**. Amsterdam: North-Holland, 1997. P. 1433–1456. Cit. on p. 46.

OMAIR, Badr; ALTURKI, Ahmad. An improved method for taxonomy development in information systems. **International Journal of Advanced Computer Science and**

Applications, v. 11, n. 4, p. 535–540, 2020. DOI: 10.14569/ijacsa.2020.0110470. Cit. on p. 50.

OMORONYIA, Inah et al. A review of awareness in distributed collaborative software engineering. English. **Software - Practice and Experience**, v. 40, n. 12, p. 1107–1133, 2010. ISSN 0038-0644. DOI: 10.1002/spe.1005. Cit. on pp. 82, 273–275, 277, 278, 280.

OSBORNE, Jason W. **Best practices in exploratory factor analysis**. Scotts Valley, CA, US: CreateSpace Independent Publishing Platform, 2014. Cit. on pp. 71, 192, 232.

OULASVIRTA, Antti. Designing Mobile Awareness Cues. In: PROCEEDINGS of the 10th international conference on Human computer interaction with mobile devices and services. Amsterdam, The Netherlands: Association for Computing Machinery, 2008. (MobileHCI '08). DOI: 10.1145/1409240.1409246. Cit. on pp. 88, 167.

PALMER, Thomas D; FIELDS, N Ann. Computer supported cooperative work. **Computer**, IEEE, v. 27, n. 5, p. 15–17, 1994. DOI: 10.1109/2.291295. Cit. on pp. 46, 47.

PASQUALI, Luiz. **TRI-Teoria de resposta ao item: Teoria, procedimentos e aplicacoes**. Curitiba, PR, BR: Editora Appris, 2020. Cit. on pp. 12, 13, 43, 130, 184.

PASQUALI, Luiz; PRIMI, Ricardo. Fundamentos da teoria da resposta ao item: TRI. **Avaliação Psicológica: Interamerican Journal of Psychological Assessment**, Instituto Brasileiro de Avaliação Psicológica (IBAP), v. 2, n. 2, p. 99–110, 2003. Cit. on pp. 12–14, 44, 53, 58, 69, 70, 72, 73, 135, 136, 193, 196, 201, 233, 236, 241.

PELEGRINA, Ana Belen et al. Integrating groupware applications into shared workspaces. English. In: 2010 4th International Conference on Research Challenges in Information Science - Proceedings, RCIS 2010. Nice, France: IEEE, 2010. P. 557–568. DOI: 10.1109/rcis.2010.5507305. Cit. on p. 82.

PETERSEN, Kai; VAKKALANKA, Sairam; KUZNIARZ, Ludwik. Guidelines for conducting systematic mapping studies in software engineering: An update. English. **Information and Software Technology**, v. 64, p. 1–18, 2015. ISSN 0950-5849. DOI: 10.1016/j.infsof.2015.03.007. Cit. on pp. 42, 77, 89, 96, 98, 242.

PETTICREW, Mark; ROBERTS, Helen. **Systematic Reviews in the Social Sciences: A Practical Guide**. Main Street, Malden, MA, USA: Blackwell Publishing Ltd, Jan. 2006. DOI: 10.1002/9780470754887. Cit. on pp. 10, 42, 77, 78, 98, 242.

POLIT, Denise F.; BECK, Cheryl Tatano. Generalization in quantitative and qualitative research: Myths and strategies. **International Journal of Nursing Studies**, Elsevier BV, v. 47, n. 11, p. 1451–1458, Nov. 2010. ISSN 0020-7489. DOI: 10.1016/j.ijnurstu.2010.06.004. Cit. on pp. 15, 244, 245.

POULOVASSILIS, Alex; XHAFA, Fatos. Building event-based services for awareness in P2P groupware systems. English. In: PROCEEDINGS - 2013 8th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing. Compiegne, France: ACM, 2013. (3PGCIC 2013), p. 200–207. DOI: 10.1109/3pgcic.2013.36. Cit. on pp. 82, 91, 93, 94, 273–275, 277, 280.

POULOVASSILIS, Alex; XHAFA, Fatos; O'HAGAN, Thomas. Event-Based Awareness Services for P2P Groupware Systems. English. **Informatica (Netherlands)**, v. 26, n. 1, p. 135–157, 2015. ISSN 0868-4952. DOI: 10.15388/informatica.2015.42. Cit. on pp. 82, 91, 93, 275, 277, 279, 280.

PRAT, Nicolas; COMYN-WATTIAU, Isabelle; AKOKA, Jacky. A taxonomy of evaluation methods for information systems artifacts. **Journal of Management Information Systems**, Taylor & Francis, v. 32, n. 3, p. 229–267, 2015. DOI: 10.1080/07421222.2015.1099390. Cit. on p. 50.

PROUZEAU, Arnaud; BEZERIANOS, Anastasia; CHAPUIS, Olivier. Awareness Techniques to Aid Transitions between Personal and Shared Workspaces in Multi-Display Environments. English. In: ISS 2018 - Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces. Tokyo, Japan: ACM, 2018. P. 291–304. DOI: 10.1145/3279778.3279780. Cit. on pp. 36, 39, 73, 75, 82, 90, 94, 95, 274, 280.

RABIN, Laura A. et al. Linking self-perceived cognitive functioning questionnaires using item response theory: The subjective cognitive decline initiative. **Neuropsychology**, American Psychological Association (APA), v. 37, n. 4, p. 463–499, May 2023. ISSN 0894-4105. DOI: 10.1037/neu0000888. Cit. on p. 53.

RECKASE, Mark D. Unifactor latent trait models applied to multifactor tests: Results and implications. **Journal of educational statistics**, Sage Publications Sage CA:

Thousand Oaks, CA, v. 4, n. 3, p. 207–230, 1979. DOI: 10.2307/1164671. Cit. on pp. 193, 235.

REINHARDT, Wolfgang et al. Understanding the meaning of awareness in research networks. English. In: CEUR Workshop Proceedings. Saarbrucken, Germany: CEUR-WS.org, 2012. P. 13–29. Cit. on pp. 82, 273–275, 280.

RICHARDSON, Roberto Jarry. **Pesquisa social: Métodos e técnicas**. São Paulo, SP, BR: Editora Atlas SA, Oct. 2017. Cit. on pp. 69, 135, 138.

RITTENBRUCH, Markus; MCEWAN, Gregor. An Historical Reflection of Awareness in Collaboration. In: **Awareness Systems: Advances in Theory, Methodology and Design**. Ed. by Panos Markopoulos, Boris De Ruyter and Wendy Mackay. London: Springer London, 2009. P. 3–48. ISBN 978-1-84882-477-5. DOI: 10.1007/978-1-84882-477-5_1. Cit. on p. 49.

RITTGEN, Peter. Quality and Perceived Usefulness of Process Models. In: (SAC '10), p. 65–72. DOI: 10.1145/1774088.1774105. Cit. on pp. 113, 114.

ROCKER, Carsten. Universal access to awareness information: Using smart artefacts to mediate awareness in distributed teams. English. **Universal Access in the Information Society**, v. 11, n. 3, p. 259–271, 2012. ISSN 1615-5289. Cit. on pp. 82, 91, 93–95.

ROGERS, Yvonne; SHARP, Helen; PREECE, Jennifer. **Design de interação**. Porto Alegre, RS, BR: Bookman Editora, 2013. Cit. on p. 117.

ROSQVIST, Tony. **Software Quality Journal**, Springer Science and Business Media LLC, v. 11, n. 1, p. 39–55, 2003. ISSN 0963-9314. DOI: 10.1023/a:1023741528816. Cit. on p. 138.

SAMEJIMA, Fumiko. Estimation of latent ability using a response pattern of graded scores. **Psychometrika monograph supplement**, 1969. DOI: 10.1007/bf03372160. Cit. on pp. 14, 56–58, 61, 62, 128–130, 201, 209, 241, 244.

SÁNCHEZ-GORDÓN, Mary; COLOMO-PALACIOS, Ricardo. Taking the emotional pulse of software engineering - A systematic literature review of empirical studies. **Information and Software Technology**, v. 115, p. 23–43, 2019. ISSN 0950-5849. DOI: 10.1016/j.infsof.2019.08.002. Cit. on p. 96.

SANSIVIERI, Valentina; WIBERG, Marie; MATTEUCCI, Mariagiulia. A review of test equating methods with a special focus on IRT-based approaches. **Statistica**, v. 77, n. 4, p. 329–352, 2017. Cit. on pp. 65, 67, 68.

SANTOS, Kathyana Vanessa Diniz; ARANTES, Fernanda Paes, et al. Supplier's performance evaluation in supply chains using item response theory. **International Journal of Services and Operations Management**, Inderscience Publishers, v. 45, n. 4, p. 530–551, 2023. ISSN 1744-2389. DOI: 10.1504/ijsom.2023.132858. Cit. on p. 53.

SANTOS, Natalia Sales; FERREIRA, Lidia Silva; PRATES, Raquel Oliveira. An Overview of Evaluation Methods for Collaborative Systems. In: 2012 Brazilian Symposium on Collaborative Systems. São Paulo, SP, BR: IEEE, Oct. 2012. DOI: 10.1109/sbsc.2012.29. Cit. on pp. 9, 36.

SCHMETTOW, Martin; VIETZE, Wolfgang. Introducing item response theory for measuring usability inspection processes. In: PROCEEDINGS of the SIGCHI Conference on Human Factors in Computing Systems. Florence, Italy: Association for Computing Machinery, Apr. 2008. (CHI '08). DOI: 10.1145/1357054.1357196. Cit. on p. 53.

SCHMIDT, Kjeld; RANDALL, Dave. Preface to the Special Issue on 'Reconsidering "Awareness" in CSCW'. **Computer Supported Cooperative Work (CSCW)**, v. 25, n. 4, p. 229–233, Oct. 2016. ISSN 1573-7551. DOI: 10.1007/s10606-016-9257-6. Cit. on pp. 35, 90.

SCHMITT, Neal. Uses and abuses of coefficient alpha. **Psychological assessment**, American Psychological Association, v. 8, n. 4, p. 350, 1996. DOI: 10.1037/1040-3590.8.4.350. Cit. on pp. 71, 192.

SCHUMACKER, Randall E. Item response theory. **Applied Measurement Associates**, 2005. Cit. on p. 54.

SCOTT, Terry F.; SCHUMAYER, Daniel. Students' proficiency scores within multitrait item response theory. **Physical Review Special Topics - Physics Education Research**, American Physical Society (APS), v. 11, n. 2, Nov. 2015. ISSN 1554-9178. DOI: 10.1103/physrevstper.11.020134. Cit. on p. 53.

SEDOC, João; UNGAR, Lyle. Item Response Theory for Efficient Human Evaluation of Chatbots. In: EGER, Steffen et al. (Eds.). **Proceedings of the First Workshop on**

Evaluation and Comparison of NLP Systems. Online: Association for Computational Linguistics, Nov. 2020. P. 21–33. DOI: 10.18653/v1/2020.eval4nlp-1.3. Cit. on p. 53.

SEEBACH, Christoph; BECK, Roman; PAHLKE, Immanuel. Situation awareness through social collaboration platforms in distributed work environments. English. In: INTERNATIONAL Conference on Information Systems 2011, ICIS 2011. Shanghai, China: Association for Information Systems, 2011. P. 1–22. Cit. on pp. 33, 34, 49, 82.

SINGH, Jagdip. Tackling measurement problems with Item Response Theory: Principles, characteristics, and assessment, with an illustrative example. **Journal of Business Research**, Elsevier, v. 57, n. 2, p. 184–208, 2004. Cit. on pp. 71, 192.

SOEGAARD, Mads; DAM, Rikke Friis. The encyclopedia of human-computer interaction. **The encyclopedia of human-computer interaction**, The Interaction Design Foundation, 2012. Cit. on p. 46.

SOUZA, Luiz Gustavo; BARBOSA, Simone Diniz Junqueira. Evaluating the MoLICC Notation Using the Cognitive Dimensions of Notations Framework. In: PROCEEDINGS of the 14th Brazilian Symposium on Human Factors in Computing Systems. Salvador, BA, BR: Association for Computing Machinery, 2015. (IHC '15), p. 1–10. DOI: 10.1145/3148456.3148474. Cit. on pp. 73–75, 82, 90–92, 94, 102.

SPENCER, Steven G. **The strength of multidimensional item response theory in exploring construct space that is multidimensional and correlated**. Provo, UT, USA: Brigham Young University, 2004. Cit. on pp. 71, 192.

STEINMACHER, Igor; CHAVES, Ana Paula; GEROSA, Marco Aurelio. Awareness Support in Distributed Software Development: A Systematic Review and Mapping of the Literature. **Computer Supported Cooperative Work (CSCW)**, Kluwer Academic Publishers, USA, v. 22, n. 2-3, 2013. ISSN 0925-9724. DOI: 10.1007/s10606-012-9164-4. Cit. on pp. 9, 36, 74, 98.

STEINMACHER, Igor; GEROSA, Marco; CHAVES, Ana Paula. Awareness Support in Global Software Development: A Systematic Review Based on the 3C Collaboration Model. In: COLLABORATION and Technology. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. P. 185–201. DOI: 10.1007/978-3-642-15714-1_15. Cit. on pp. 47, 74.

STERNBERG, R.J.; MIO, J.S. **Cognitive psychology**. 5. ed. Belmount, CA, USA: Cengage Learning, 2006. Cit. on p. 48.

STERNBERG, Robert J; STERNBERG, Karin; MIO, JS. **Cognitive Psychology**. Belmont, California, USA: Wadsworth: Cengage Learning, 2012. P. 167–169. Cit. on p. 34.

STEVENS, Stanley Smith. Handbook of experimental psychology. Wiley, 1951. DOI: 10.2307/1418844. Cit. on pp. 51, 52.

STRAETMANS, Gerard J. J. M.; EGGEN, Theo J. H. M. Computerized Adaptive Testing: What It Is and How It Works. **Educational Technology**, Educational Technology Publications, Inc., v. 38, n. 1, p. 45–52, 2024/05/15/ 1998. ISSN 00131962. Cit. on p. 246.

STRASSER, Artur. Delphi method variants in information systems research: Taxonomy development and application. **Electronic Journal of Business Research Methods**, v. 15, n. 2, p. 120–133, 2017. Cit. on pp. 51, 113, 114, 138, 142.

SZOPINSKI, Daniel; SCHOORMANN, Thorsten; KUNDISCH, Dennis. Because Your Taxonomy is Worth IT: towards a Framework for Taxonomy Evaluation. In: 27TH European Conference on Information Systems. Stockholm University, Kista, Sweden: Association for Information Systems, 2019. Cit. on pp. 10, 42, 43, 50, 51, 102, 104, 203, 243.

TALAEI-KHOEI, Amir; RAY, Pradeep, et al. A framework for awareness maintenance. English. **Journal of Network and Computer Applications**, v. 35, n. 1, p. 199–210, 2012. ISSN 1084-8045. DOI: 10.1016/j.jnca.2011.06.011. Cit. on pp. 82, 91, 93.

TALAEI-KHOEI, Amir; VICHITVANICHPHONG, Suchada, et al. A methodology to develop awareness in computer supported collaborative work using policies. English. **Journal of Computer and System Sciences**, v. 80, n. 7, p. 1323–1338, 2014. ISSN 0022-0000. DOI: 10.1016/j.jcss.2014.03.003. Cit. on pp. 73, 82, 90, 275–277, 280.

TANG, Tiffany Y; WINOTO, Pinata; LEUNG, Hareton. A usability study of an educational groupware system: Supporting awareness for collaboration. **Journal of Educational Computing Research**, SAGE Publications Sage CA: Los Angeles, CA, v. 50, n. 3, p. 379–402, 2014. DOI: 10.2190/ec.50.3.e. Cit. on p. 49.

TAVAKOL, Mohsen; WETZEL, Angela. Factor Analysis: a means for theory and instrument development in support of construct validity. en. **Int J Med Educ**, England, UK, v. 11, p. 245–247, Nov. 2020. DOI: 10.5116/ijme.5f96.0f4a. Cit. on pp. 14, 15, 196, 201, 236, 241, 244.

TENENBERG, Josh; ROTH, Wolff-Michael; SOCHA, David. From I-Awareness to We-Awareness in CSCW. **Computer Supported Cooperative Work (CSCW)**, v. 25, n. 4, p. 235–278, Oct. 2016. ISSN 1573-7551. DOI: 10.1007/s10606-014-9215-0. Cit. on pp. 8, 33.

TERUEL, Miguel A. et al. A comprehensive framework for modeling requirements of CSCW systems. English. **Journal of Software: Evolution and Process**, Elsevier, v. 29, n. 5, p. 1–19, 2017. ISSN 2047-7481. DOI: 10.1002/smr.1858. Cit. on pp. 82, 91, 92, 94.

TEZZA, Rafael; BORNIA, Antonio Cezar; ANDRADE, Dalton Francisco de. Measuring web usability using item response theory: Principles, features and opportunities. **Interacting with Computers**, v. 23, n. 2, p. 167–175, 2011. ISSN 0953-5438. DOI: 10.1016/j.intcom.2011.02.004. Cit. on p. 53.

TEZZA, Rafael; BORNIA, Antonio Cezar; ANDRADE, Dalton Francisco de; BARBETTA, Pedro Alberto. Modelo multidimensional para mensurar qualidade em website de e-commerce utilizando a teoria da resposta ao item. **Gestão e Produção**, FapUNIFESP (SciELO), v. 25, n. 4, p. 916–934, Dec. 2018. ISSN 0104-530X. DOI: 10.1590/0104-530x1875-18. Cit. on p. 53.

THISSEN, David; WAINER, Howard. **Test Scoring**. New York, NY, USA: Routledge, 2001. DOI: 10.1111/j.1745-3984.2002.tb01178.x. Cit. on pp. 158, 164, 183, 191, 224, 232.

THOMAS, Godwin; BOTHA, Reinhardt A.; VAN GREUNEN, Darrell. Towards addressing the requirements for coordination in the distributed South African public sector from a virtual community-centric collaboration process lifecycle perspective. English. In: 2015 IST-Africa Conference. Lilongwe, Malawi: IST-Africa, 2015. P. 1–11. DOI: 10.1109/istafrica.2015.7190581. Cit. on p. 82.

THOMPSON, Bruce. Exploratory and confirmatory factor analysis: Understanding concepts and applications. **Washington, DC, USA**, Citeseer, v. 10694, n. 000, p. 3, 2004. Cit. on pp. 72, 192, 232.

TRIERWEILLER, Andréa Cristina et al. Measuring organizational effectiveness in information and communication technology companies using item response theory. **Work**, IOS Press, v. 41, p. 2795–2802, 2012. ISSN 1051-9815. DOI: 10.3233/wor-2012-0526-2795. Cit. on p. 53.

TRIPATHI, Anand. A transaction model with multilevel consistency for shared data in distributed groupware systems. English. In: PROCEEDINGS - 2016 IEEE 2nd International Conference on Collaboration and Internet Computing, IEEE CIC 2016. Pittsburgh, PA, United states: IEEE, 2016. P. 236–245. DOI: 10.1109/cic.2016.041. Cit. on pp. 82, 275–277, 280.

TROCHIM, William MK; DONNELLY, James P. **Research methods knowledge base**. New York, NY, USA: Atomic Dog Pub. Macmillan Publishing Company, 2001. v. 2. Cit. on pp. 12, 13, 44, 136, 243.

URBINA, Susana. **Fundamentos da testagem psicológica**. Porto Alegre, RS, BR: Artmed Editora, 2009. Cit. on pp. 70, 135.

USMAN, Muhammad et al. Taxonomies in software engineering: A Systematic mapping study and a revised taxonomy development method. English. **Information and Software Technology**, v. 85, p. 43–59, 2017. ISSN 0950-5849. DOI: 10.1016/j.infsof.2017.01.006. Cit. on pp. 10, 42, 43, 77, 102, 104, 243.

VAN SOLINGEN, Rini; BERGHOUT, Egon W. **The Goal/Question/Metric Method: a practical guide for quality improvement of software development**. Berkshire, UK: McGraw-Hill International (UK) Limited, 1999. Cit. on pp. 43, 139.

VENABLE, John; PRIES-HEJE, Jan; BASKERVILLE, Richard. A comprehensive framework for evaluation in design science research. In: DESIGN Science Research in Information Systems. Advances in Theory and Practice: 7th International Conference, DESRIST 2012. Las Vegas, NV, USA: Springer, May 2012. P. 423–438. Cit. on p. 40.

VINCENZI, Silvana Ligia et al. Assessment of environmental sustainability perception through item response theory: A case study in Brazil. **Journal of Cleaner Production**, Elsevier BV, v. 170, p. 1369–1386, Jan. 2018. ISSN 0959-6526. DOI: 10.1016/j.jclepro.2017.09.217. Cit. on p. 53.

WEISS, David J. Improving Measurement Quality and Efficiency with Adaptive Testing. **Applied Psychological Measurement**, SAGE Publications, v. 6, n. 4, p. 473–492, Sept. 1982. ISSN 1552-3497. DOI: 10.1177/014662168200600408. Available from: <http://dx.doi.org/10.1177/014662168200600408>. Cit. on p. 246.

WEISS, DJ. Polychotomous or polytomous? **Applied psychological measurement**, v. 19, n. 1, 1995. DOI: 10.1177/014662169501900102. Cit. on p. 56.

WIERINGA, Roel et al. Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. **Requirements Engineering**, v. 11, n. 1, p. 102–107, Mar. 2006. ISSN 1432-010X. DOI: 10.1007/s00766-005-0021-6. Cit. on p. 89.

WIRTH, R. J.; EDWARDS, Michael C. Item factor analysis: Current approaches and future directions. **Psychological Methods**, American Psychological Association (APA), v. 12, n. 1, p. 58–79, Mar. 2007. ISSN 1082-989X. DOI: 10.1037/1082-989x.12.1.58. Available from: <http://dx.doi.org/10.1037/1082-989X.12.1.58>. Cit. on pp. 146, 165, 202.

WOHLIN, Claes. 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. London, England, United Kingdom: Association for Computing Machinery, 2014. (EASE '14), p. 1–10. DOI: 10.1145/2601248.2601268. Cit. on pp. 79, 98, 122.

WOHLIN, Claes et al. **Experimentation in software engineering: an introduction**. New York, NY, USA: Springer Science & Business Media, 2012. DOI: 10.1007/978-3-642-29044-2. Cit. on pp. 12, 13, 44, 122, 135, 243.

XHAFA, Fatos et al. Coordination in android mobile teams. English. In: PROCEEDINGS - 16th International Conference on Network-Based Information Systems, NBiS 2013. Gwangju, Republic of Korea: IEEE, 2013. P. 482–487. DOI: 10.1109/nbis.2013.80. Cit. on pp. 82, 91, 93, 94, 272, 273, 276, 277.

YANG, Chi-Lan et al. Knowing That You Know What I Know Helps? Understanding the Effects of Knowledge Transparency in Online Knowledge Transfer. English. **ACM Transaction on Computer-Human Interaction**, Association for Computing Machinery, New York, NY, USA, v. 2, CSCW, p. 1–21, Nov. 2018. DOI: 10.1145/3274458. Cit. on pp. 73, 82, 90, 91, 274.

YANG, Lanxin et al. Quality Assessment in Systematic Literature Reviews: A Software Engineering Perspective. **Information and Software Technology**, v. 1, n. 1, p. 1–24, 2020. ISSN 0950-5849. DOI: 10.1016/j.infsof.2020.106397. Cit. on p. 96.

YIN, Robert K. **Case study research: Design and methods**. London, UK: Sage Publications, Inc, 2009. v. 5. Cit. on pp. 13, 44, 135, 243.

YUILL, Nicola; ROGERS, Yvonne. Mechanisms for collaboration: A design and evaluation framework for multi-user interfaces. English. **ACM Transactions on Computer-Human Interaction**, v. 19, n. 1, p. 1–25, 2012. ISSN 1073-0516. DOI: 10.1145/2147783.2147784. Cit. on pp. 82, 92, 276.

Appendix

APPENDIX A – SYSTEMATIC MAPPING MATERIALS

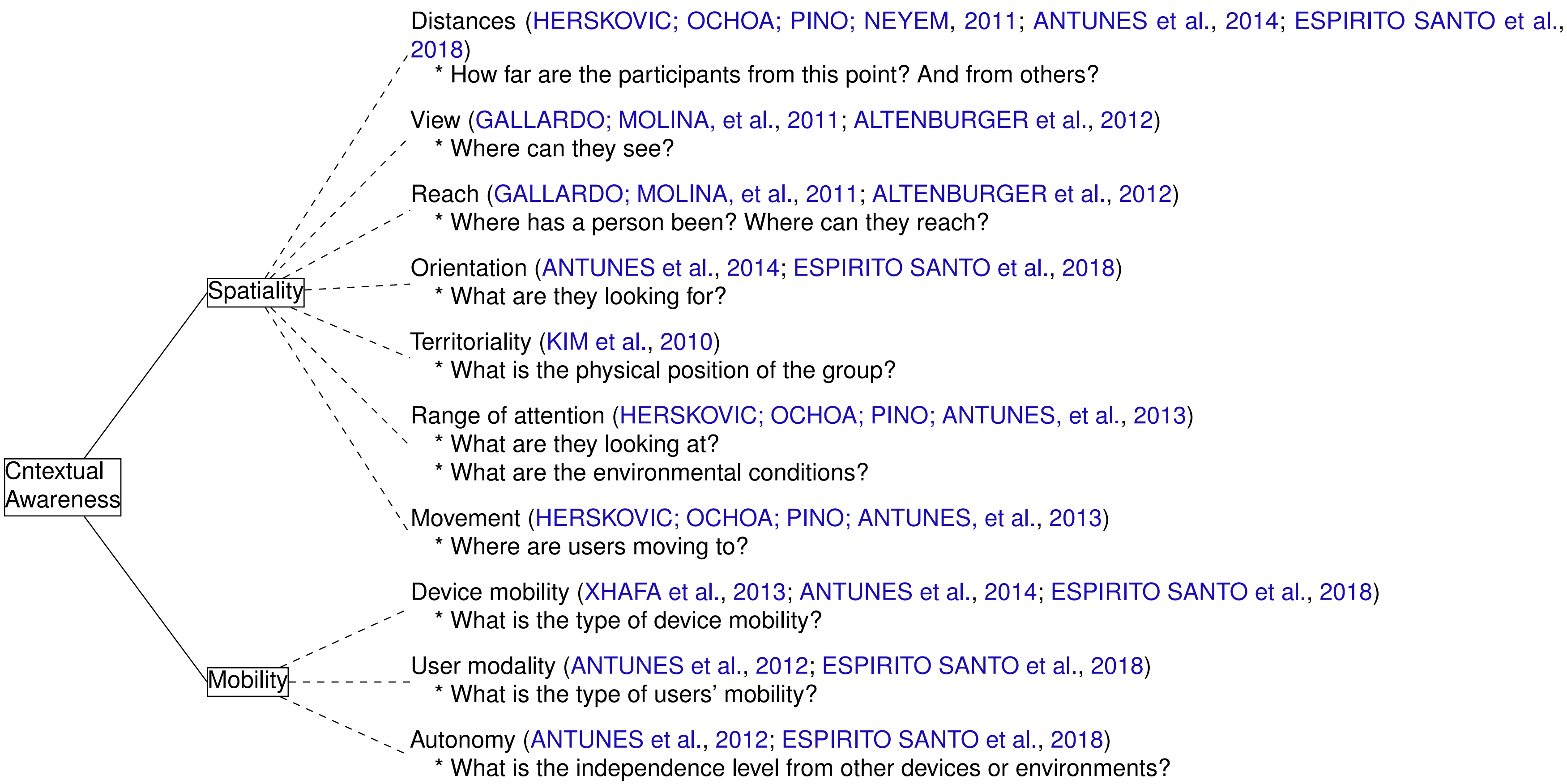


Figure 59 – Contextual Awareness view (part 1 of 2)

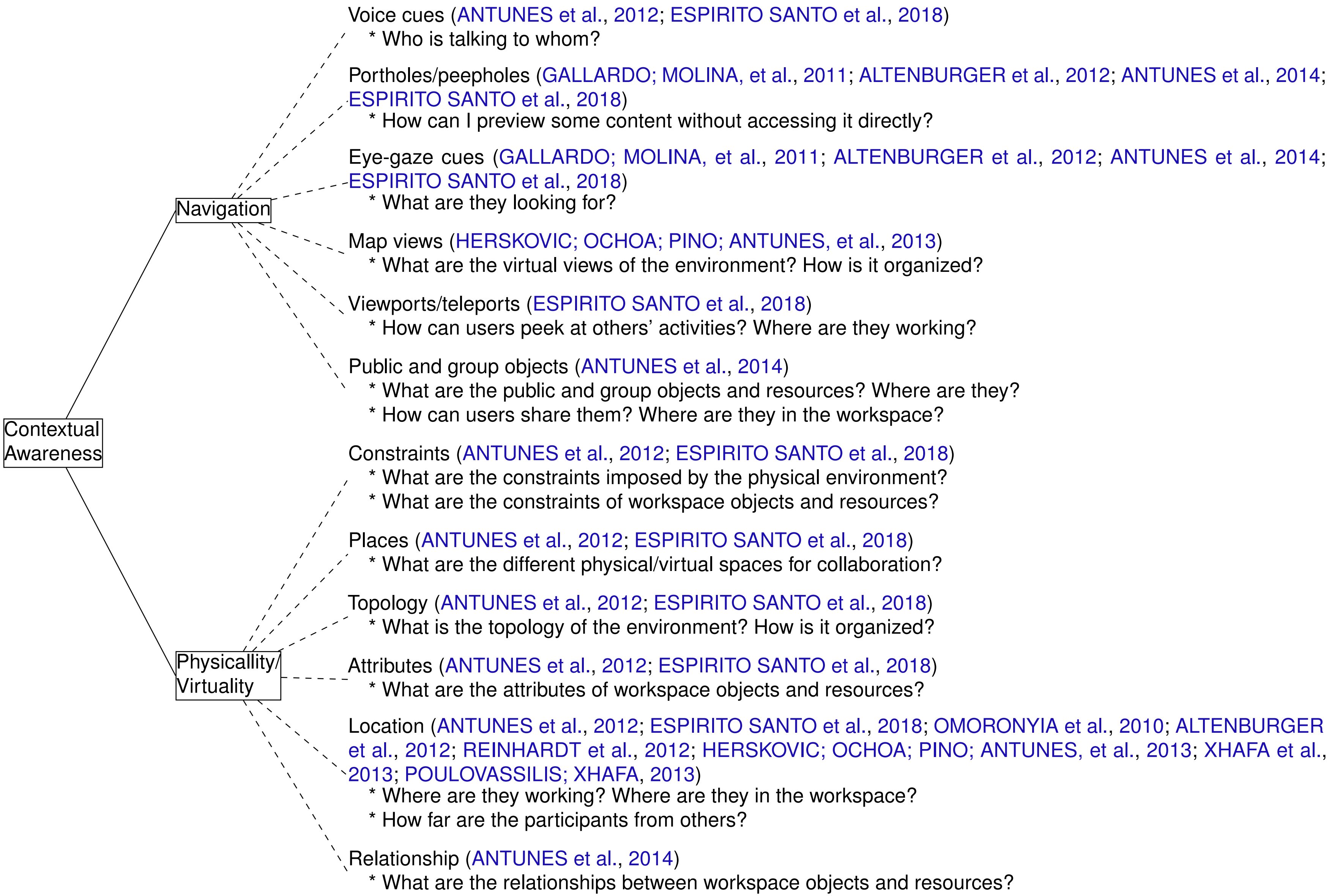


Figure 60 – Contextual Awareness view (part 2 of 2)

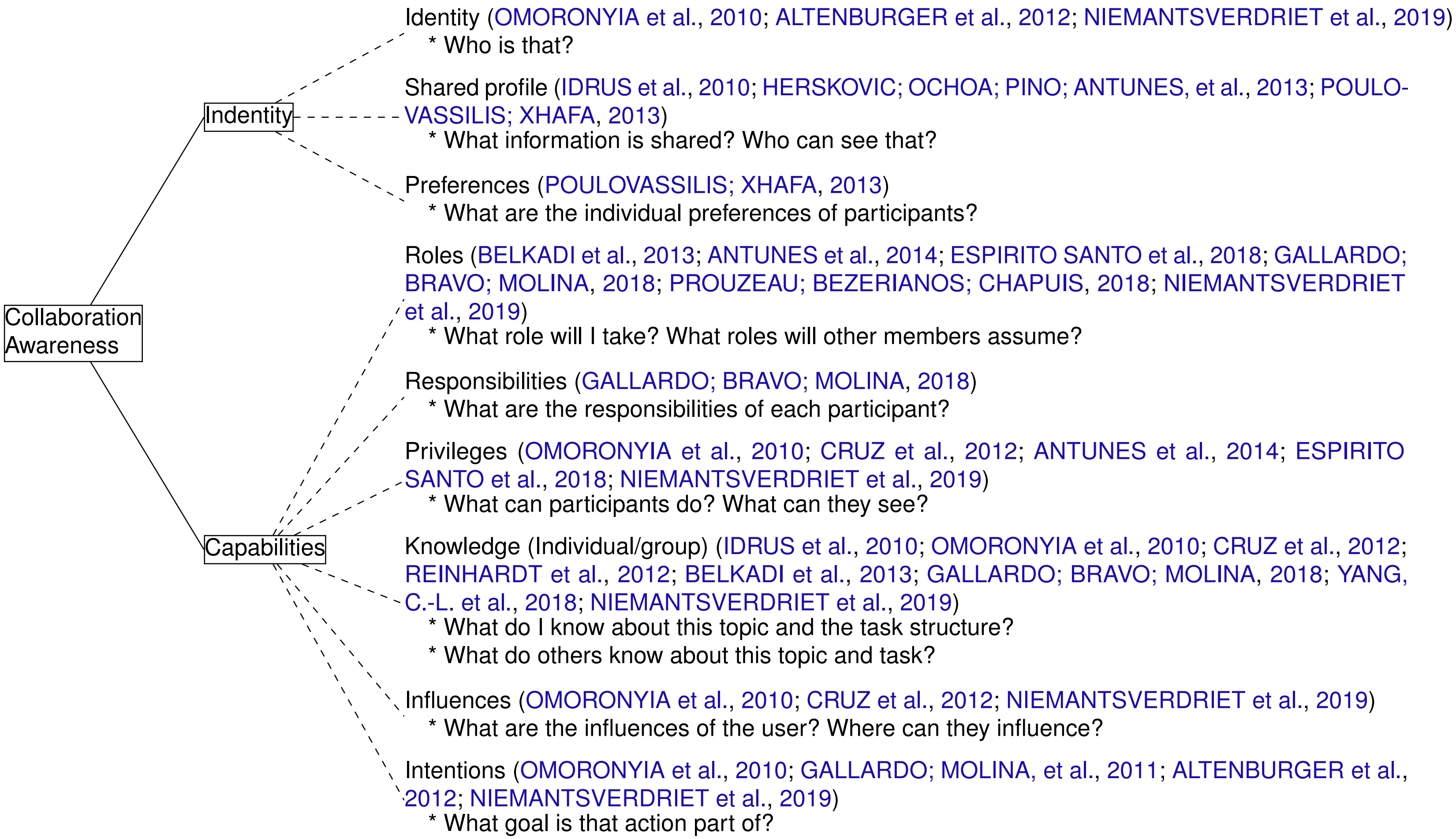


Figure 61 – Collaboration Awareness view (part 1 of 2)

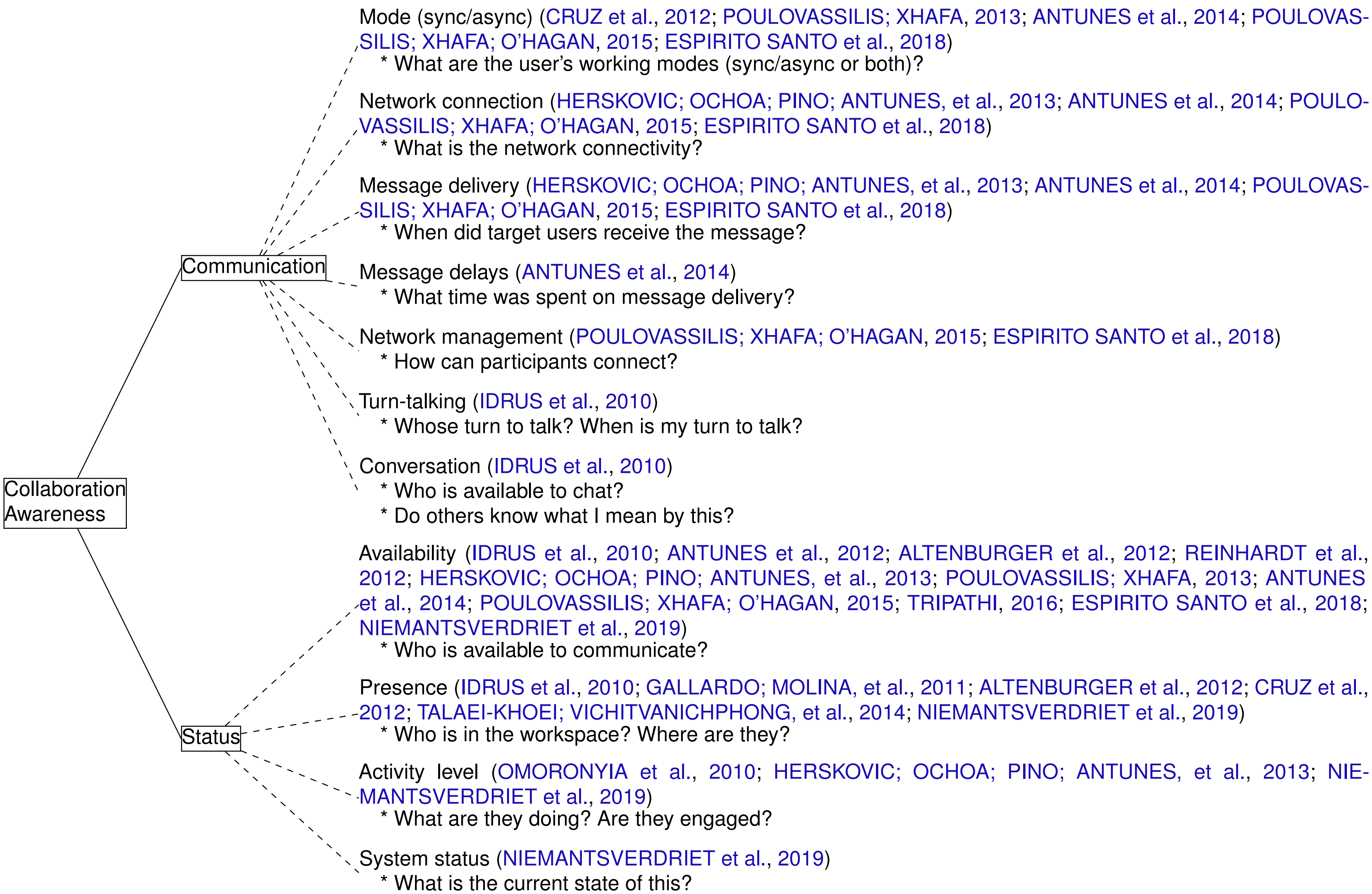


Figure 62 – Collaboration Awareness view (part 2 of 2)

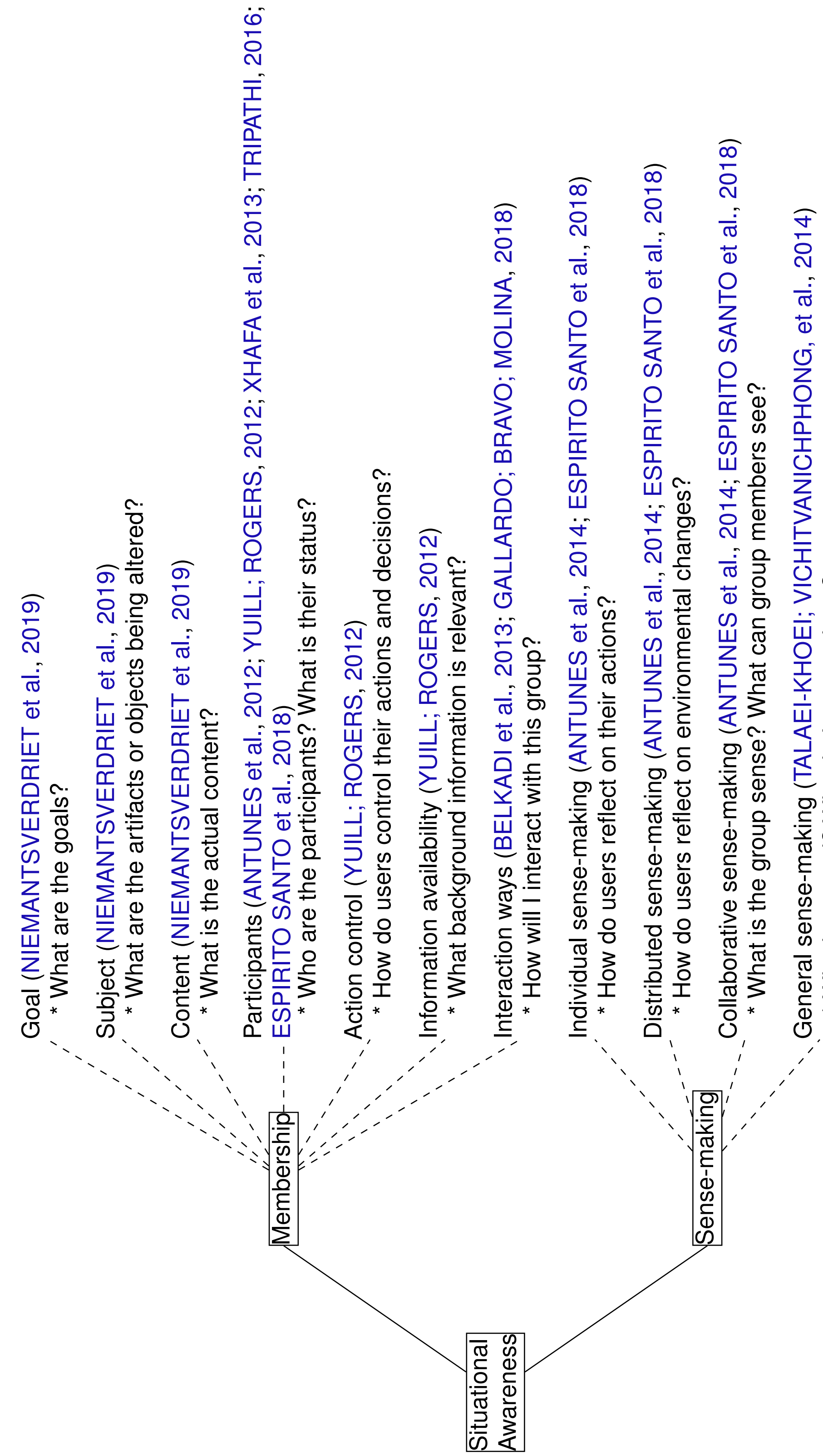


Figure 63 – Situational Awareness view (part 1 of 2)

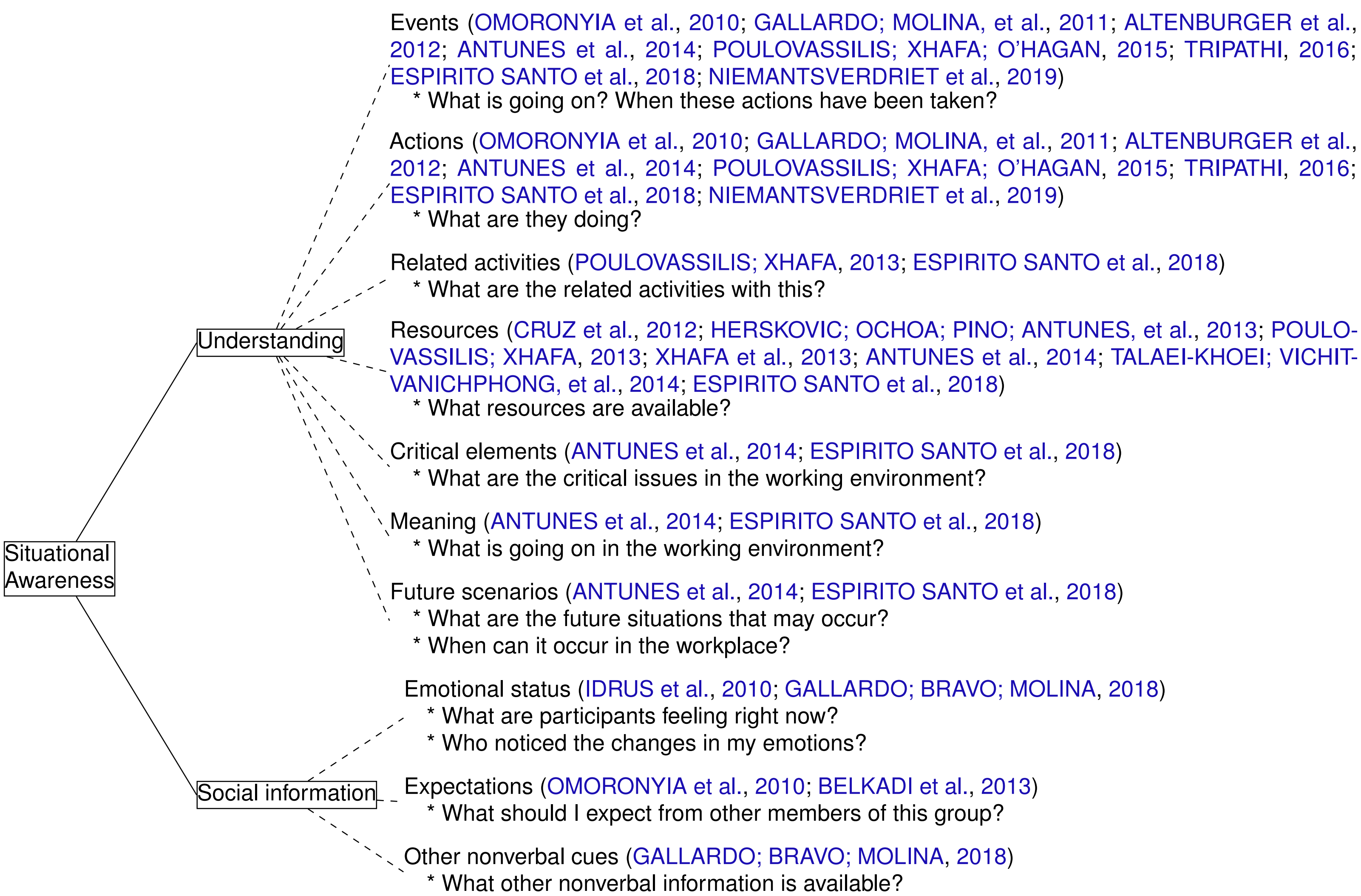


Figure 64 – Situational Awareness view (part 2 of 2)

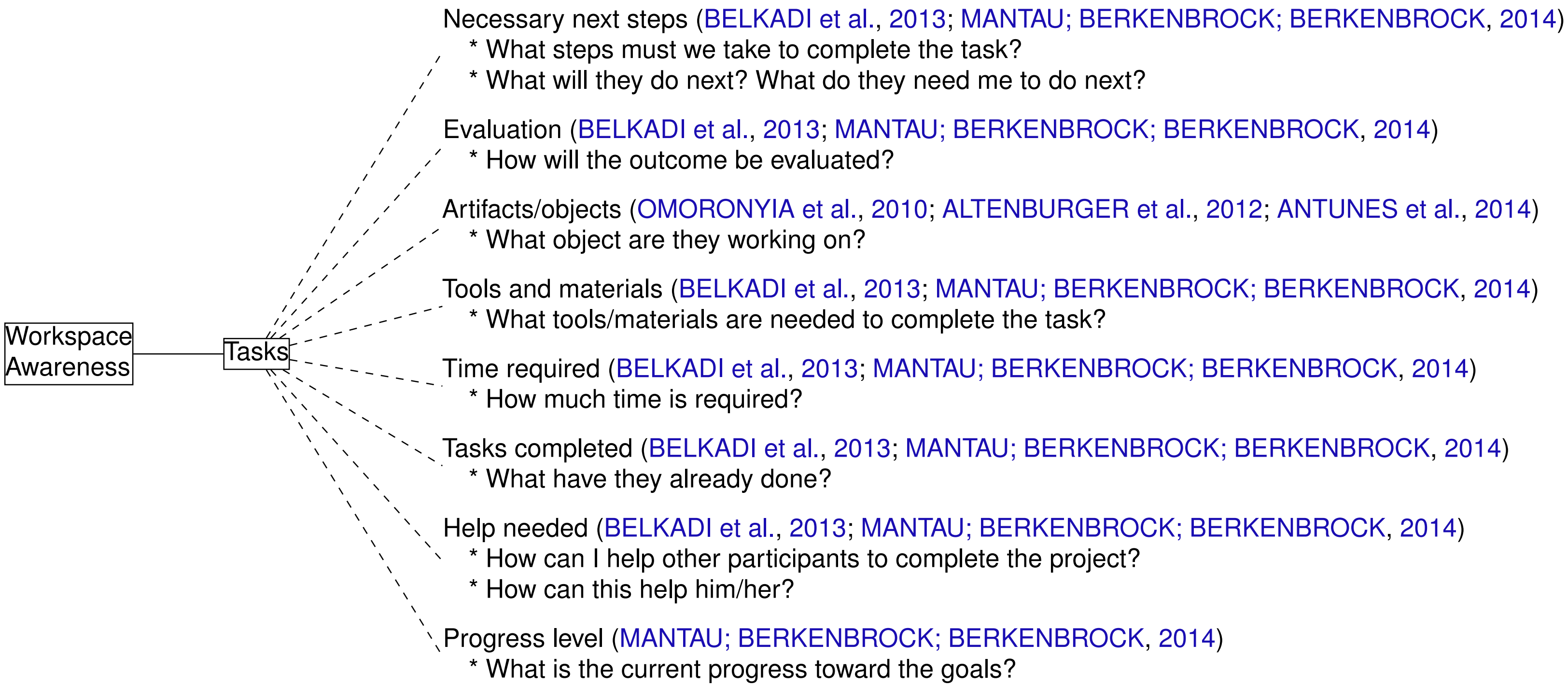


Figure 65 – Workspace Awareness view (part 1 of 2)

Figure 66 – Workspace Awareness view (part 2 of 2)

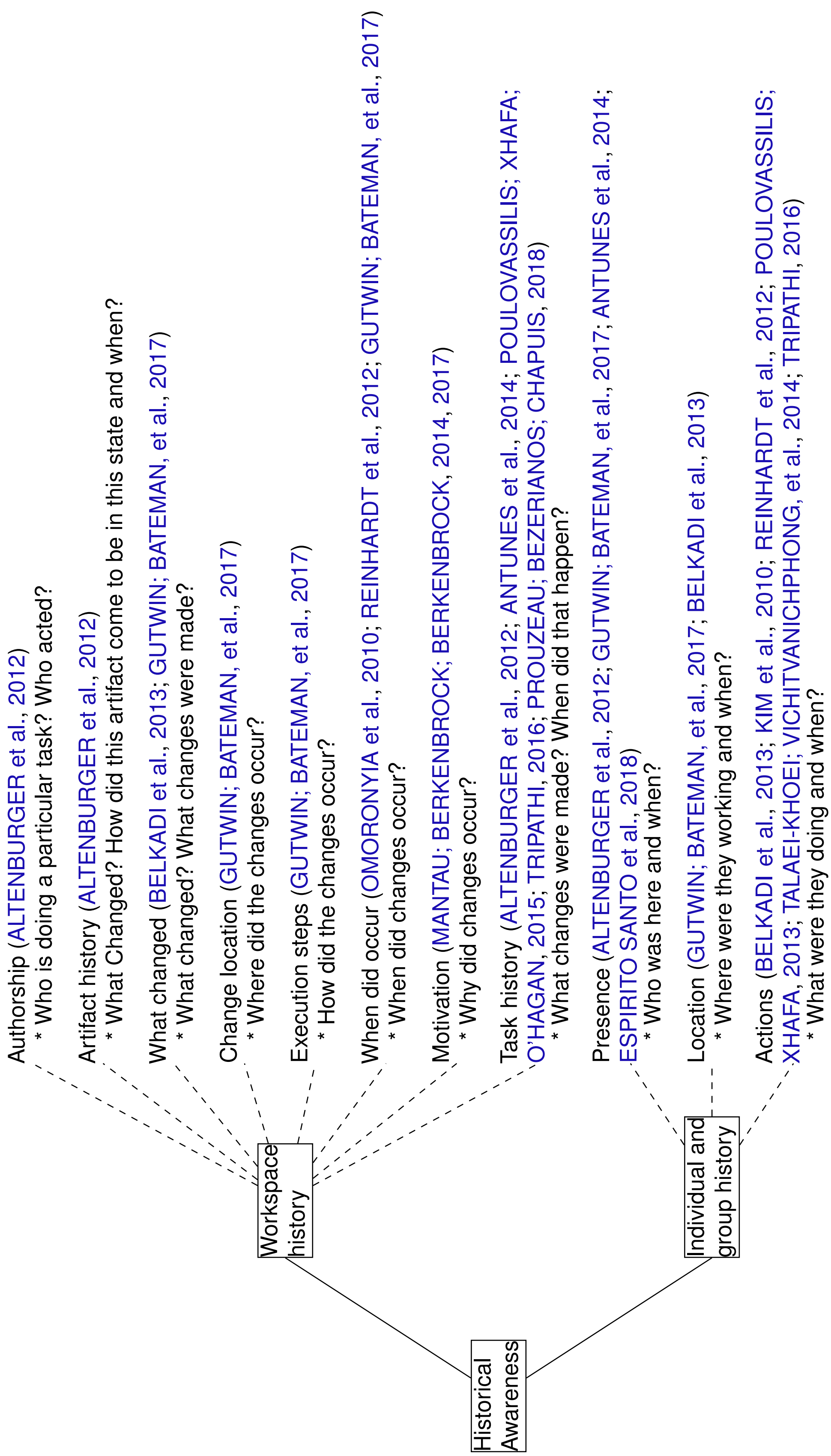


Figure 67 – Historical Awareness view

APPENDIX B – AWARENESS TAXONOMY MATERIALS

Table 53 – Collaboration awareness category

	Design element	5W+1H Classification
Identity	Identity	(who) – Who are the participants?
	Shared profile	(what/who) – What information is shared? Who can see that?
	Preferences	(what) – What are the individual preferences of participants?
Capabilities	Roles	(what) – What role will I take in this group? What roles will the other members of the group assume?
	Responsibilities	(what) – What are the responsibilities of each participant?
	Privileges	(what) – What can participants do? What can they see?
	Knowledge	(what) – What do I know about this topic and the task structure? What do others know?
	Influences	(what/where) – What are the influences of the user and where?
	Intentions	(what) – What goal is that action part of?
Status	Availability	(who/how) – Who can communicate and how? What background information is relevant to users' behavior and tasks?
	Presence	(who/where) – Who are in the workspace? Where are they? Who was here and when?
	Activity level	(what) – What are they doing? Are they engaged?
	Status	(what) – What are their status?
Communication	Mode	(what) – What are the user's working modes (online, offline, or both)?
	Message connectivity	(what) – What is the network connectivity?
	Message delivery	(when) – When did the target users receive the messages?
	Message delays	(what) – What time was spent in message delivery?
	Interactions ways	(what) – How can participants connect? How will I interact with this group?
	Turn-talking	(who/when) – Whose turn to talk? When is my turn?
	Conversation	(who/what) – Do others know what I mean by this?
Social	Expectations	(what) – What should I expect from other group members?
	Emotional status	(what) – What are participants feeling right now? Does anyone notice the changes in my emotions?
	Nonverbal cues	(what) – What other nonverbal information is available?

Table 54 – Workspace awareness category

	Design element	5W+1H Classification
Activities	Goal	(what) – What are the goals?
	Subject	(what) – What are the artifacts or objects being altered?
	Content	(what) – What is the actual content?
	Motivation	(why) – Why did changes occur?
	Time required	(how) – How much time is required?
	Progress level	(what) – What is the current progress towards the goals? What have they already done?
	Help needed	(how) – How can I help other participants to complete the project?
	Evaluation	(how) – How will the outcome be evaluated?
Workflow	Authorship	(who) – Who is doing the task? Who acted?
	Execution steps	(what/how/when) – What steps must we take to complete the task? How did the changes occur? When did that event happen? What changed? What is the activity performed by a particular user?
	Events and actions	(what/when) – What is going on? When these actions have been taken? What were they doing individually/as a group, and when?
	Change location	(where) – Where did the changes occur?
	Related activities	(what) – What are the related activities with this?
	Parallel activities	(what/who) – What activities are doing parallel? Who is doing them?
	Coordinated activities	(what/who) – What activities are coordinated? Who is doing them?
	Mutually adjusted activities	(what/who) – What activities are mutually adjusted? Who is doing them?
Environment	Tools and materials	(what) – What tools/materials are needed to complete the task?
	Artifacts and objects	(what/how/when) – What object are they working on? How did this artifact come to be in this state, and when? What changes were made? When did that happen?
	Resources availability	(what) – What resources are available?
	Critical elements	(what) – What are the critical issues in the working environment?

continues on the next page

Table 54: Workspace awareness category (continuation)

	Virtual relationships	(what) – What are the relationships between objects and resources in the workspace?
Understanding	Meaning	(what) – What is happening in the working environment?
	Scenarios	(what) – What future situations may occur? When can it occur in the workplace?
	Sensemaking	(who/what/how) Individual – How do users reflect on their actions? Distributed – How do users reflect on environmental changes? Collaborative – What is the group sense? What do group members see their goals and achieve? General – Who is around, and what belongs to them?
Relationship	Action control	(how) – How do users control their actions and decisions?
	Access control	(who/how) – Who controls a shared object or resource, and how?
	Access privileges	(who/what) – What are the privileges of the data or group activities? Who can access them?
	Control mechanisms	(what) – What control mechanisms are being used?
Interaction	Feedback	(what) – What are the responses to actions performed by the user?
	Feedthrough	(who/what) – What are the responses to actions performed by other users? Who carried out these actions?
	Backchannel feedback	(who/how) – Who is following his/her actions? How can this help him/her?
	Feedforward	(what/where) – What effects are propagated by ongoing tasks? Where did the changes take place?

Table 55 – Contextual awareness category

	Design element	5W+1H Classification
Spatiallity	Location	(Where/when/how) – Where were they working and when? What are they going to do? What are they doing? What are their current activities and tasks? Where are they physically? What is the physical position of the group? How far are the participants physically from others? What are the other members of the group doing to complete the task?
	Distances	(how) – How far are the other participants from this point?

continues on the next page

Table 55: Contextual awareness category (continuation)

	Constraints	(what) – What are the constraints imposed by the physical environment where it is used? What is the workspace constraint about objects and resources?
	Places	(what/where/how) – How is the support regarding physical places? What are the different places for collaboration? Where can participants collaborate?
	Topology	(what/where) – What is the complexity of the physical environment where it is used? What is the topology of the environment? What is the topology of the virtual environment? What are the virtual spaces? Where are they in the shared workspace?
	Attributes	(what) – What are the environmental conditions where it is used? What are the attributes of objects and resources in the workspace?
	View	(where) – Where can they see?
	Reach	(where) – Where has a person been? Where can they reach?
	Orientation	(where) – Where are they moving toward?
	Movement	(where) – Where are users moving to?
	Range of attention	(what) – What are they looking at?
Mobility	User modality	(what) – What are the types of users' modalities?
	User mobility	(what) – What is the type of device mobility?
	Autonomy	(what) – What is the level of independence from other devices or the environment?
Navigation	Voice cues	(who) – Who is talking to whom?
	Portholes/peepholes	(how) – How can I preview some content without accessing it directly?
	Eye-gaze cues	(where) – Where are they looking?
	Map views	(what/where/how) – What are the virtual views of the environment? How is it organized? Where are they?
	Viewports/Teleports	(where/how) – How can users peek the others' activities? Where are they working?
	Objects/Artifacts location	(what/where/how) – What are the objects and resources? How can users share them? Where are they in the workspace?

Table 56 – Collaboration awareness view assessment

Our taxonomy		Antunes et al. (2014)	Espirito Santo et al. (2018)	Gallardo, Bravo, and Molina (2018)	Niemantsverdriet et al. (2019)
Identity	Identity	–	Participants	Who→identity	Actors→identity
	Shared profile	–	–	–	–
	Preferences	–	–	–	–
Capabilities	Roles	Practice→roles	Membership→roles	Group structure→roles	Capabilities→roles
	Responsibilities	–	–	Group structure→ responsibilities	Capabilities→abilities
	Privileges	Background→privileges	Membership→privileges	–	–
	Knowledge	–	–	Informal information→ knowledge	Capabilities→skills
	Influences	–	–	–	Capabilities→influences
	Intentions	–	–	What→intention	Capabilities→Intentions, plans
Status	Availability	Availability	Accessibility→place (same, any, different, co-located, virtual, remote)	–	Status→availability
	Presence	–	–	Who→presence, presence history	Status→presence
	Activity level	–	–	–	Status→activity level
	System status	–	–	–	Situation→system status
Communication	Mode	Communication mode	Sync/Async	–	–
	Network connectivity	Network connectivity	Network connectivity	–	–
	Message delivery	Message delivery	Message delivery	–	–
	Message delays	Message delays	–	–	–
	Interaction ways	–	Network management	–	–
	Turn-talking	–	–	–	–
	Conversation	–	–	–	–
Social	Expectations	–	–	–	–
	Emotional status	–	–	Emotional status	–
	Nonverbal cues	–	–	Other nonverbal cues	–

Table 57 – Workspace awareness view assessment

Our taxonomy		Antunes et al. (2014)	Espirito Santo et al. (2018)	Gallardo, Bravo, and Molina (2018)	Niemantsverdriet et al. (2019)
Activities	Goal	–	–	–	Interaction→goal
	Subject	–	–	–	Interaction→subject
	Content	–	–	–	Interaction→content
	Motivation	–	–	–	–
	Time required	–	–	–	–
	Progress level	–	–	–	–
	Help needed	–	–	–	–
	Evaluation	–	–	–	–
Workflow	Authorship	Who?	Who?	Who→authorship	Interaction→authorship
	Execution steps	What? How? When?	What? How? When?	When→event history	–
	Events and actions	Events/Actions	Events/Actions	Who→action history	Interaction→events
	Change location	Where?	Where?	–	–
	Related activities	–	–	–	–
	Parallel activities	Parallel activities	Parallel activities	–	–
	Coordinated activities	Coordinated activities	Coordinated activities	–	–
	Mutually adjusted activities	Mutually adjusted activities	Mutually adjusted activities	–	–
Understanding	Tolls and materials	–	–	–	–
	Artifacts and objects	–	–	What→artifact	–
	Resource availability	Resources	Resources	–	–
	Critical elements	Critical elements	Critical elements	–	–
	Virtual relationships	Virtual relationship	Virtual relationship	–	–
	Meaning	Meaning	Meaning	–	–
	Scenarios	Future Scenarios	Future Scenarios	–	–
	Sense-making	Sense-making (individual, distributed, collaborative)	Sense-making (individual, distributed, collaborative)	–	–
Interaction	Feedback	Feedback	Feedback	–	–
	Feedthrough	Feedthrough	Feedthrough	–	–
	Backchannel feedback	Backchannel feedback	Backchannel feedback	–	–
	Feedforward	–	–	–	–
Relationship	Action control	–	–	–	–
	Access control	Access control	Access control	–	–
	Access privileges	–	–	–	–
	Control mechanisms	–	–	–	–

Table 58 – Contextual awareness view assessment

Our taxonomy		Antunes et al. (2014)	Espirito Santo et al. (2018)	Gallardo, Bravo, and Molina (2018)	Niemantsverdriet et al. (2019)
Location	User location	Cartesian/Topological locations	Cartesian/Topological	Who→location history/ Where→location	–
	Distances	Distances	Distances	–	–
	Constraints	Physical constraints/ Virtual constraints	Physical constraints/ Virtual constraints	–	–
	Places	Physical places/ Navigation→virtual places	Physical places/ Virtual places	–	–
	Topology	Physical topology/ Navigation→virtual topology	Physical topology/ Virtual topology	–	–
	Attributes	Physical attributes/ Virtual attributes	Physical attributes/ Virtual attributes	–	–
	View	–	–	–	–
	Reach	–	–	–	–
	Orientation	Orientation	Orientation	–	–
	Movement	–	–	–	–
	Range of attention	Range of attention	Range of attention	–	Actors→attention level
Mobility	User modality	Location modality	–	–	–
	User mobility	Level of mobility	Mobility→Wandering/Visiting	–	–
	Autonomy	Relation with other devices	–	–	–
Navigation	Voice cues	Interaction→voice cues	–	–	–
	Portholes/Peepholes	–	–	–	–
	Eye-gaze cues	Interaction→eye-gaze cues	–	Where→gaze	–
	Map views	Map views	–	Where→view	–
	Viewports/Teleports	Viewport/Teleport	Viewport/Teleport	Where→reach	–
	Objects/Artifacts location	Virtuality→group objects/ Virtuality→public objects	Virtuality→group/ Virtuality→public	–	–

APPENDIX C – ASSESSMENT MODEL MATERIALS (PORTUGUESE VERSION ONLY)

Avaliação do Suporte à Colaboração na Plataforma Moodle

Este instrumento busca avaliar os aspectos de colaboração fornecidos na plataforma Moodle. O Moodle é uma plataforma de aprendizagem colaborativa projetada para criar ambientes de aprendizagem, destinada a educadores e alunos. Nestes ambientes virtuais de aprendizagem, para que a colaboração ocorra de forma satisfatória, é desejável que o ambiente forneça aos participantes mecanismos de suporte à percepção.

Por meio deste instrumento, buscamos identificar **o nível de suporte à colaboração fornecido pelo ambiente colaborativo** Moodle sob a perspectiva do usuário, seja **estudante, educador, ou administrador.** Ao responder as questões aqui apresentadas, considere sempre o conjunto de **recursos do Moodle que você comumente utilizada, a exemplo, chat, fórum, tarefas, lição, wiki.**

O modelo de avaliação para ambientes colaborativos utilizado permite avaliar o nível de suporte à colaboração sob a perspectiva do participante. O instrumento está dividido em duas partes: Parte 1 – Formulário demográfico (6 questões); Parte 2 – Inventário da percepção (30 questões de suporte à percepção). Estima-se de 5 a 10 minutos para responder o questionário.

Todas as informações são coletadas de forma anônima e voluntária, utilizados somente para a avaliação do ambiente colaborativo objeto do estudo. **Asseguramos o anonimato das informações fornecidas**, em todas as etapas da pesquisa.

Obrigado por sua participação!

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Você está sendo convidado(a) para participar, como voluntário(a), em uma pesquisa científica.

Este TCLE se refere ao projeto de pesquisa de doutorado "*An Awareness Assessment Model for Ccollaborative Interfaces*", cujo objetivo é "desenvolver um modelo de avaliação para ambientes colaborativos por meio do suporte aos mecanismos de percepção fornecidos". Para ter uma cópia deste TCLE você deverá imprimi-lo, ou deverá gerar uma cópia em PDF para guardá-lo em seu computador. Você também poderá solicitar aos pesquisadores do estudo uma versão deste documento a qualquer momento por um dos e-mails registrados no final deste termo.

A pesquisa será realizada por meio de um questionário online e/ou impresso, constituído por duas partes: Parte 1 – Formulário demográfico (6 questões); Parte 2 – Inventário da percepção (30 questões de suporte à percepção). Estima-se que você precisará entre 5 a 10 minutos para responder o questionário. A precisão de suas respostas é determinante para a qualidade da pesquisa. O risco da pesquisa é mínimo por envolver apenas a resposta ao questionário online e/ou impresso, o qual foi elaborado com o intuito de que o tempo gasto para seu preenchimento seja mínimo. Para garantir a confidencialidade e a privacidade dos indivíduos, a caracterização dos mesmos será feita por codificação de sua identidade. Todos os dados obtidos na pesquisa serão utilizados exclusivamente com finalidades científicas conforme previsto no consentimento do participante. Os resultados da pesquisa não serão divulgados a terceiros.

Caso você não queira participar, não há problema algum. Você não precisa me explicar porque, e não haverá nenhum tipo de punição por isso. Você tem todo o direito de não querer participar do estudo, basta selecionar a opção correspondente no final desta página. O(a) senhor(a) poderá se retirar do estudo a qualquer momento, sem qualquer necessidade de justificativa. Asseguramos o anonimato das informações fornecidas, em todas as etapas da pesquisa. Não existe benefício ou vantagem direta em participar deste estudo. Os benefícios e vantagens em participar são indiretos, proporcionando retorno social através de melhorias nos ambientes colaborativos avaliados e da publicação dos resultados da pesquisa em periódicos científicos. As pessoas que acompanharão os procedimentos serão os pesquisadores: Msc. Márcio J. Mantau e/ou Dra. Fabiane B.V. Benitti, que são os responsáveis pela pesquisa.

Informe-nos caso tenha dúvidas, sugestões, comentários sobre o instrumento. Ficaremos felizes em receber feedback e ajudá-lo se necessário. Se preferir, escreva para Márcio J. Mantau (marcio.mantau@udesc.br) ou Fabiane B.V. Benitti (fabiane.benitti@ufsc.br).

Eu concordo em participar voluntariamente do presente estudo como participante, conforme termos apresentados no Termo de Consentimento Livre e Esclarecido (TCLE).

Foi me informado sobre tudo o que vai acontecer na pesquisa, o que terei que fazer, inclusive sobre os possíveis riscos e benefícios envolvidos na minha participação. O pesquisador me garantiu que eu poderei sair da pesquisa a qualquer momento, sem dar nenhuma explicação, e que esta decisão não me trará nenhum tipo de penalidade ou interrupção de meu tratamento. Fui informado também que devo imprimir ou gerar um pdf do TCLE para ter uma cópia e que posso solicitar uma versão dele via e-mail para os pesquisadores.

Ao prosseguir, aceito participar voluntariamente da presente pesquisa.

PARTE 1 – FORMULÁRIO DEMOGRÁFICO

Q1. Ambiente colaborativo avaliado: [____]

Q2. Gênero (opcional): [ ] Masculino [ ] Feminino [ ] Outro: [________________________]

Q4. Faixa etária: [ ] até 17 anos [ ] 18 a 28 anos [ ] 29 a 39 anos [ ] 40 a 50 anos [ ] acima de 50 anos

Julgue os seguintes itens, considerando a escala: de 1(novato) a 4 (experiente).

Q5. De forma geral, qual é o seu nível de familiaridade com o ambiente avaliado?	① ② ③ ④
Q6. De forma geral, qual é sua experiência no uso de ambientes colaborativos?	① ② ③ ④
Q7. De forma geral, qual é sua experiência nos conceitos de sistemas colaborativos e percepção?	① ② ③ ④

Figure 68 – Awareness Assessment Model Questionnaire: Moodle example

Avaliação do Suporte à Colaboração em Ambientes Virtuais (videoconferência)

Este instrumento busca avaliar ambientes virtuais de colaboração, projetados para a comunicação/interação entre duas ou mais pessoas de forma simultânea. A exemplo, ambientes de conferência, videoconferência, eventos virtuais, webinars, etc. **Responda as questões considerando o ambiente virtual de colaboração (videoconferência) de sua preferência.**

Para que haja colaboração, ambientes virtuais devem fornecer aos participantes mecanismos de suporte à percepção. Estes mecanismos são informações, elementos ou pistas disponíveis no ambiente colaborativo que permitem a compreensão do ambiente, das atividades, dos demais envolvidos, e do contexto para ações realizadas.

O modelo de avaliação para ambientes colaborativos utilizado permite avaliar o nível de suporte à colaboração sob a perspectiva do participante. O instrumento está dividido em duas partes: Parte 1 – Formulário demográfico (6 questões); Parte 2 – Inventário da percepção (30 questões de suporte à percepção). Estima-se de 5 a 10 minutos para responder o questionário.

Todas as informações são coletadas de forma anônima e voluntária, utilizados somente para a avaliação do ambiente colaborativo objeto do estudo. **Asseguramos o anonimato das informações fornecidas**, em todas as etapas da pesquisa.

Obrigado por sua participação!

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Você está sendo convidado(a) para participar, como voluntário(a), em uma pesquisa científica.

Este TCLE se refere ao projeto de pesquisa de doutorado "*An Awareness Assessment Model for Ccollaborative Interfaces*", cujo objetivo é "desenvolver um modelo de avaliação para ambientes colaborativos por meio do suporte aos mecanismos de percepção fornecidos". Para ter uma cópia deste TCLE você deverá imprimi-lo, ou deverá gerar uma cópia em PDF para guardá-lo em seu computador. Você também poderá solicitar aos pesquisadores do estudo uma versão deste documento a qualquer momento por um dos e-mails registrados no final deste termo.

A pesquisa será realizada por meio de um questionário online e/ou impresso, constituído por duas partes: Parte 1 – Formulário demográfico (6 questões); Parte 2 – Inventário da percepção (30 questões de suporte à percepção). Estima-se que você precisará entre 5 a 10 minutos para responder o questionário. A precisão de suas respostas é determinante para a qualidade da pesquisa. O risco da pesquisa é mínimo por envolver apenas a resposta ao questionário online e/ou impresso, o qual foi elaborado com o intuito de que o tempo gasto para seu preenchimento seja mínimo. Para garantir a confidencialidade e a privacidade dos indivíduos, a caracterização dos mesmos será feita por codificação de sua identidade. Todos os dados obtidos na pesquisa serão utilizados exclusivamente com finalidades científicas conforme previsto no consentimento do participante. Os resultados da pesquisa não serão divulgados a terceiros.

Caso você não queira participar, não há problema algum. Você não precisa me explicar porque, e não haverá nenhum tipo de punição por isso. Você tem todo o direito de não querer participar do estudo, basta selecionar a opção correspondente no final desta página. O(a) senhor(a) poderá se retirar do estudo a qualquer momento, sem qualquer necessidade de justificativa. Asseguramos o anonimato das informações fornecidas, em todas as etapas da pesquisa. Não existe benefício ou vantagem direta em participar deste estudo. Os benefícios e vantagens em participar são indiretos, proporcionando retorno social através de melhorias nos ambientes colaborativos avaliados e da publicação dos resultados da pesquisa em periódicos científicos. As pessoas que acompanharão os procedimentos serão os pesquisadores: Msc. Márcio J. Mantau e/ou Dra. Fabiane B.V. Benitti, que são os responsáveis pela pesquisa.

Informe-nos caso tenha dúvidas, sugestões, comentários sobre o instrumento. Ficaremos felizes em receber feedback e ajudá-lo se necessário. Se preferir, escreva para Márcio J. Mantau (marcio.mantau@udesc.br) ou Fabiane B.V. Benitti (fabiane.benitti@ufsc.br).

Eu concordo em participar voluntariamente do presente estudo como participante, conforme termos apresentados no Termo de Consentimento Livre e Esclarecido (TCLE).

Foi me informado sobre tudo o que vai acontecer na pesquisa, o que terei que fazer, inclusive sobre os possíveis riscos e benefícios envolvidos na minha participação. O pesquisador me garantiu que eu poderei sair da pesquisa a qualquer momento, sem dar nenhuma explicação, e que esta decisão não me trará nenhum tipo de penalidade ou interrupção de meu tratamento. Fui informado também que devo imprimir ou gerar um pdf do TCLE para ter uma cópia e que posso solicitar uma versão dele via e-mail para os pesquisadores.

Ao prosseguir, aceito participar voluntariamente da presente pesquisa.

PARTE 1 – FORMULÁRIO DEMOGRÁFICO

Q1. Ambiente colaborativo avaliado: [____]

Q2. Gênero (opcional): **[ ]** Masculino [ ] Feminino [ ] Outro: [__________________________]

Q4. Faixa etária: [ ] até 17 anos [ ] 18 a 28 anos [ ] 29 a 39 anos [ ] 40 a 50 anos [ ] acima de 50 anos

Julgue os seguintes itens, considerando a escala: de 1(novato) a 4 (experiente).

Q5. De forma geral, qual é o seu nível de familiaridade com o ambiente avaliado?	① ② ③ ④
Q6. De forma geral, qual é sua experiência no uso de ambientes colaborativos?	① ② ③ ④
Q7. De forma geral, qual é sua experiência nos conceitos de sistemas colaborativos e percepção?	① ② ③ ④

Figure 69 – Awareness Assessment Model Questionnaire: Videoconferencing example

PARTE 2 – INVENTÁRIO DA PERCEPÇÃO	
Julgue os seguintes itens, considerando a escala: ① Discordo totalmente; ② Discordo; ③ Concordo; e ④ Concordo totalmente.	
Q1. O ambiente colaborativo me permite identificar o objetivo das tarefas/atividades realizadas.	① ② ③ ④
Q2. Quando estou interagindo, eu consigo identificar quais são os artefatos ou objetos que estão sendo alterados.	① ② ③ ④
Q3. Quando estou interagindo, eu consigo acompanhar o progresso da tarefa/atividade realizada em grupo.	① ② ③ ④
Q4. Ao terminar uma tarefa/atividade conjunta, eu consigo avaliar os resultados obtidos.	① ② ③ ④
Q5. O ambiente colaborativo me permite identificar quem está conduzindo as tarefas/atividades (autoria).	① ② ③ ④
Q6. Durante a interação, eu consigo identificar se os outros participantes estão engajados com a tarefa/atividade atual ou se estão realizando outras tarefas/atividades em paralelo.	① ② ③ ④
Q7. Ao utilizar o ambiente colaborativo, eu consigo identificar as ferramentas e materiais disponíveis para colaborar.	① ② ③ ④
Q8. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem restrições/elementos críticos para a realização das tarefas/atividades.	① ② ③ ④
Q9. Quando estou realizando uma atividade em grupo, eu consigo identificar a compreensão dos demais participantes envolvidos.	① ② ③ ④
Q10. Quando estou realizando uma atividade em grupo, eu consigo acompanhar as alterações realizadas pelos demais participantes.	① ② ③ ④
Q11. Ao utilizar o ambiente colaborativo, eu consigo identificar quem está controlando o ambiente, as tarefas/atividades, ou os recursos compartilhados.	① ② ③ ④
Q12. Ao utilizar o ambiente colaborativo, eu consigo visualizar a identidade dos participantes (quem são?).	① ② ③ ④
Q13. Quando estou utilizando o ambiente colaborativo, eu consigo identificar quais informações estão sendo compartilhadas.	① ② ③ ④
Q14. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem regras distintas (e quais são) para cada participante.	① ② ③ ④
Q15. Durante a interação, eu consigo identificar o que cada participante pode fazer, ver, ou ainda, controlar.	① ② ③ ④
Q16. Quando estou interagindo, eu consigo identificar o que eu sei sobre a tarefa/atividade atual e como eu posso ajudar.	① ② ③ ④
Q17. Ao utilizar o ambiente colaborativo, eu consigo identificar a disponibilidade de cada participante.	① ② ③ ④
Q18. Ao utilizar o ambiente colaborativo, eu consigo identificar a presença de cada participante no ambiente.	① ② ③ ④
Q19. Ao utilizar o ambiente colaborativo, eu consigo identificar o modo de trabalho (síncrono ou assíncrono).	① ② ③ ④
Q20. Durante a interação, sou notificado se ocorrem atrasos no envio/recebimento das mensagens.	① ② ③ ④
Q21. Durante a interação, eu consigo identificar quem está conversando, trocando ideias, ou de quem é a vez de falar.	① ② ③ ④
Q22. Durante a interação, eu consigo identificar quais são as expectativas envolvendo cada participante.	① ② ③ ④
Q23. Ao utilizar o ambiente colaborativo, eu consigo visualizar a localização física/virtual dos demais participantes.	① ② ③ ④
Q24. Ao utilizar o ambiente colaborativo, eu consigo identificar como o ambiente está organizado (topologia).	① ② ③ ④
Q25. O ambiente colaborativo me permite identificar a orientação/direção de cada participante.	① ② ③ ④
Q26. Durante a interação, eu consigo acompanhar a movimentação de cada participante no ambiente compartilhado.	① ② ③ ④
Q27. Durante a interação, eu consigo identificar o nível de atenção necessário para a realização da tarefa/atividade.	① ② ③ ④
Q28. O ambiente colaborativo me permite identificar se o sistema permite diferentes modalidades/dispositivos de acesso (local/remoto, parado/movimento, etc.).	① ② ③ ④
Q29. O ambiente colaborativo possibilita certa mobilidade ao usuário (acesso por diferentes dispositivos).	① ② ③ ④
Q30. Ao utilizar o ambiente colaborativo, eu consigo identificar se existe uma dependência entre a aplicação e o local de utilização (recursos/funcionalidades disponíveis depende do local de acesso).	① ② ③ ④

Caderno 1 = {01, 02, 06, 08, 09, 14, 17, 20, 24, 28, 30, 33, 34, 36, 38, 39, 42, 43, 46, 49, 51, 53, 56, 60, 64, 65, 66, 67, 68, 69}

Figure 70 – Awareness Assessment Model Questionnaire: book 1

PARTE 2 – INVENTÁRIO DA PERCEPÇÃO	
Julgue os seguintes itens, considerando a escala: ① Discordo totalmente; ② Discordo; ③ Concordo; e ④ Concordo totalmente.	
Q1. Ao utilizar o ambiente colaborativo, eu consigo acessar o conteúdo atual dos recursos compartilhados.	① ② ③ ④
Q2. Ao utilizar o ambiente colaborativo, eu consigo identificar o tempo necessário/disponível para realizar a tarefa/atividade.	① ② ③ ④
Q3. Quando estou realizando uma atividade em grupo, eu consigo compreender quais são os passos/ações que devem ser realizadas.	① ② ③ ④
Q4. Ao utilizar o ambiente colaborativo, eu consigo identificar os locais onde posso interagir (onde posso realizar as tarefas/atividades).	① ② ③ ④
Q5. Durante a interação, eu consigo identificar como a tarefa/atividade atual está relacionada ao cenário atual (tarefas/atividades conjuntas).	① ② ③ ④
Q6. Ao utilizar o ambiente colaborativo, eu consigo identificar a presença de artefatos/objetos necessários para colaborar.	① ② ③ ④
Q7. Ao utilizar o ambiente colaborativo, eu consigo identificar os recursos/funcionalidades disponíveis para colaborar.	① ② ③ ④
Q8. Ao utilizar o ambiente colaborativo, eu consigo identificar o relacionamento entre os objetos/recursos do ambiente.	① ② ③ ④
Q9. Durante a interação, eu consigo compreender o significado das ações realizadas (o que está acontecendo no ambiente).	① ② ③ ④
Q10. Durante a interação, eu consigo compreender quais são as próximas tarefas/atividades que devem ser realizadas ou situações futuras que podem ocorrer.	① ② ③ ④
Q11. Ao utilizar o ambiente colaborativo, eu consigo identificar a resposta das ações realizadas pelos demais participantes (feedthrought).	① ② ③ ④
Q12. Quando estou realizando uma atividade em grupo, eu consigo identificar se os demais estão acompanhando as ações realizadas.	① ② ③ ④
Q13. Durante a interação, eu consigo identificar como os demais participantes estão controlando suas ações/decisões.	① ② ③ ④
Q14. Ao utilizar o ambiente colaborativo, eu consigo identificar a presença de restrição de acesso (privilégios) do ambiente compartilhado (seja nas tarefas/atividades, objetos/artefatos, ou recursos).	① ② ③ ④
Q15. O ambiente colaborativo me permite identificar se existem mecanismos de controle de acesso aos recursos/funcionalidades e como posso acessá-los.	① ② ③ ④
Q16. O ambiente colaborativo me permite identificar as preferências individuais de cada participante.	① ② ③ ④
Q17. Durante a interação, eu consigo identificar as responsabilidades que cada participante pode assumir.	① ② ③ ④
Q18. Quando estou interagindo, eu consigo identificar quais são as influências/decisões de cada participante.	① ② ③ ④
Q19. Durante a interação, eu consigo expressar minhas intenções (o que pretendo fazer) no ambiente e consigo identificar as intenções dos demais participantes.	① ② ③ ④
Q20. Durante a realização das tarefas/atividades em grupo, eu consigo identificar o nível de atividade/engajamento de cada participante.	① ② ③ ④
Q21. Ao utilizar o ambiente colaborativo, eu consigo acessar a conectividade de rede.	① ② ③ ④
Q22. Durante a interação, eu consigo identificar a disponibilidade de informações não verbais para a comunicação (emojis, gestos, etc.).	① ② ③ ④
Q23. Ao utilizar o ambiente colaborativo, eu consigo visualizar a distância entre cada participante.	① ② ③ ④
Q24. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem restrições de espaço envolvidas.	① ② ③ ④
Q25. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem diferentes locais para colaboração.	① ② ③ ④
Q26. Ao utilizar o ambiente colaborativo, eu consigo identificar os atributos dos objetos/recursos ou condições do ambiente.	① ② ③ ④
Q27. O ambiente colaborativo me permite identificar o que cada participante pode ver.	① ② ③ ④
Q28. Ao utilizar o ambiente colaborativo, eu consigo identificar os meios para pré-visualizar (espiar) o conteúdo das tarefas/atividades, sem acessá-los diretamente.	① ② ③ ④
Q29. Ao utilizar o ambiente colaborativo, eu consigo identificar os meios para pré-visualizar (espiar) as tarefas/atividades realizadas pelos demais participantes.	① ② ③ ④
Q30. Ao utilizar o ambiente colaborativo, eu consigo identificar onde estão os objetos/artefatos ou recursos no ambiente compartilhado.	① ② ③ ④

Caderno 2 = {03, 05, 10, 12, 16, 18, 19, 21, 22, 23, 26, 27, 29, 31, 32, 35, 37, 40, 41, 44, 47, 55, 57, 58, 59, 61, 62, 71, 74, 75}

Figure 71 – Awareness Assessment Model Questionnaire: book 2

PARTE 2 – INVENTÁRIO DA PERCEPÇÃO	
Julgue os seguintes itens, considerando a escala: ① Discordo totalmente; ② Discordo; ③ Concordo; e ④ Concordo totalmente.	
Q1. O ambiente colaborativo me permite identificar o objetivo das tarefas/atividades realizadas.	① ② ③ ④
Q2. Quando estou interagindo, eu consigo identificar a motivação para as tarefas/atividades realizadas**.**	① ② ③ ④
Q3. Quando estou realizando uma atividade em grupo, eu consigo identificar como posso auxiliar os demais participantes.	① ② ③ ④
Q4. O ambiente colaborativo me permite identificar quem está conduzindo as tarefas/atividades (autoria).	① ② ③ ④
Q5. Durante a interação, eu consigo identificar o que está acontecendo no ambiente e o que já foi realizado.	① ② ③ ④
Q6. Durante a interação, eu consigo identificar se existem outras tarefas/atividades que estão relacionadas ao cenário atual.	① ② ③ ④
Q7. Durante a interação, eu consigo identificar se os outros participantes estão engajados com a tarefa/atividade atual ou se estão realizando outras tarefas/atividades em paralelo.	① ② ③ ④
Q8. Durante a interação, eu consigo identificar se a tarefa/atividade está sendo realizada de forma coordenada e quem está a coordenado.	① ② ③ ④
Q9. Ao utilizar o ambiente colaborativo, eu consigo identificar as ferramentas e materiais disponíveis para colaborar.	① ② ③ ④
Q10. Quando estou realizando uma atividade em grupo, eu consigo identificar a compreensão dos demais participantes envolvidos.	① ② ③ ④
Q11. Ao utilizar o ambiente colaborativo, eu consigo identificar a resposta das ações realizadas por mim (feedback).	① ② ③ ④
Q12. Ao utilizar o ambiente colaborativo, eu consigo identificar quem está controlando o ambiente, as tarefas/atividades, ou os recursos compartilhados.	① ② ③ ④
Q13. Quando estou interagindo, eu consigo identificar o que eu sei sobre a tarefa/atividade atual e como eu posso ajudar.	① ② ③ ④
Q14. Ao utilizar o ambiente colaborativo, eu consigo identificar a disponibilidade de cada participante.	① ② ③ ④
Q15. Ao utilizar o ambiente colaborativo, eu consigo identificar a presença de cada participante no ambiente.	① ② ③ ④
Q16. Ao utilizar o ambiente colaborativo, eu consigo identificar o estado/situação atual de cada participante.	① ② ③ ④
Q17. Ao utilizar o ambiente colaborativo, eu consigo identificar o modo de trabalho (síncrono ou assíncrono).	① ② ③ ④
Q18. Durante a interação, eu consigo visualizar quando as mensagens são enviadas/recebidas pelos demais participantes.	① ② ③ ④
Q19. Ao utilizar o ambiente colaborativo, eu consigo identificar os meios disponíveis para conectar e interagir com os demais participantes.	① ② ③ ④
Q20. Durante a interação, eu consigo identificar quem está conversando, trocando ideias, ou de quem é a vez de falar.	① ② ③ ④
Q21. O ambiente colaborativo me permite identificar os meios disponíveis para conversar com os demais participantes (estabelecer uma comunicação).	① ② ③ ④
Q22. Durante a interação, eu consigo identificar o estado emocional de cada participante.	① ② ③ ④
Q23. Ao utilizar o ambiente colaborativo, eu consigo visualizar a localização física/virtual dos demais participantes.	① ② ③ ④
Q24. O ambiente colaborativo me permite identificar o alcance de cada participante (onde pode ir, o que pode acessar).	① ② ③ ④
Q25. Durante a interação, eu consigo acompanhar a movimentação de cada participante no ambiente compartilhado.	① ② ③ ④
Q26. O ambiente colaborativo me permite identificar se o sistema permite diferentes modalidades/dispositivos de acesso (local/remoto, parado/movimento, etc.).	① ② ③ ④
Q27. Ao utilizar o ambiente colaborativo, eu consigo identificar se existe uma dependência entre a aplicação e o local de utilização (recursos/funcionalidades disponíveis depende do local de acesso).	① ② ③ ④
Q28. Quando estou interagindo, eu consigo identificar quem está falando com quem (comunicação verbal).	① ② ③ ④
Q29. Durante a interação, eu consigo identificar onde cada participante está olhando.	① ② ③ ④
Q30. O ambiente colaborativo me permite visualizar o ambiente compartilhado de forma simplificada (miniatura, mapa, ou outro similar).	① ② ③ ④

Caderno 3 = {01, 04, 07, 09, 11, 13, 14, 15, 17, 24, 25, 30, 39, 42, 43, 45, 46, 48, 50, 51, 52, 54, 56, 63, 65, 67, 69, 70, 72, 73}

Figure 72 – Awareness Assessment Model Questionnaire: book 3

PARTE 2 – INVENTÁRIO DA PERCEPÇÃO	
Julgue os seguintes itens, considerando a escala: ① Discordo totalmente; ② Discordo; ③ Concordo; e ④ Concordo totalmente.	
Q1. O ambiente colaborativo me permite identificar o objetivo das tarefas/atividades realizadas.	① ② ③ ④
Q2. Ao utilizar o ambiente colaborativo, eu consigo identificar o tempo necessário/disponível para realizar a tarefa/atividade.	① ② ③ ④
Q3. O ambiente colaborativo me permite identificar quem está conduzindo as tarefas/atividades (autoria).	① ② ③ ④
Q4. Quando estou realizando uma atividade em grupo, eu consigo compreender quais são os passos/ações que devem ser realizadas.	① ② ③ ④
Q5. Durante a interação, eu consigo identificar se os outros participantes estão engajados com a tarefa/atividade atual ou se estão realizando outras tarefas/atividades em paralelo.	① ② ③ ④
Q6. Durante a interação, eu consigo identificar como a tarefa/atividade atual está relacionada ao cenário atual (tarefas/atividades conjuntas).	① ② ③ ④
Q7. Ao utilizar o ambiente colaborativo, eu consigo identificar as ferramentas e materiais disponíveis para colaborar.	① ② ③ ④
Q8. Ao utilizar o ambiente colaborativo, eu consigo identificar a presença de artefatos/objetos necessários para colaborar.	① ② ③ ④
Q9. Durante a interação, eu consigo compreender o significado das ações realizadas (o que está acontecendo no ambiente).	① ② ③ ④
Q10. Quando estou realizando uma atividade em grupo, eu consigo identificar a compreensão dos demais participantes envolvidos.	① ② ③ ④
Q11. Ao utilizar o ambiente colaborativo, eu consigo identificar a resposta das ações realizadas pelos demais participantes (feedthrought).	① ② ③ ④
Q12. Ao utilizar o ambiente colaborativo, eu consigo identificar quem está controlando o ambiente, as tarefas/atividades, ou os recursos compartilhados.	① ② ③ ④
Q13. O ambiente colaborativo me permite identificar se existem mecanismos de controle de acesso aos recursos/funcionalidades e como posso acessá-los.	① ② ③ ④
Q14. O ambiente colaborativo me permite identificar as preferências individuais de cada participante.	① ② ③ ④
Q15. Durante a interação, eu consigo identificar as responsabilidades que cada participante pode assumir.	① ② ③ ④
Q16. Quando estou interagindo, eu consigo identificar o que eu sei sobre a tarefa/atividade atual e como eu posso ajudar.	① ② ③ ④
Q17. Quando estou interagindo, eu consigo identificar quais são as influências/decisões de cada participante.	① ② ③ ④
Q18. Durante a interação, eu consigo expressar minhas intenções (o que pretendo fazer) no ambiente e consigo identificar as intenções dos demais participantes.	① ② ③ ④
Q19. Ao utilizar o ambiente colaborativo, eu consigo identificar a disponibilidade de cada participante.	① ② ③ ④
Q20. Ao utilizar o ambiente colaborativo, eu consigo identificar a presença de cada participante no ambiente.	① ② ③ ④
Q21. Ao utilizar o ambiente colaborativo, eu consigo identificar o modo de trabalho (síncrono ou assíncrono).	① ② ③ ④
Q22. Durante a interação, eu consigo identificar quem está conversando, trocando ideias, ou de quem é a vez de falar.	① ② ③ ④
Q23. Ao utilizar o ambiente colaborativo, eu consigo visualizar a localização física/virtual dos demais participantes.	① ② ③ ④
Q24. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem diferentes locais para colaboração (e quais são).	① ② ③ ④
Q25. O ambiente colaborativo me permite identificar o que cada participante pode ver.	① ② ③ ④
Q26. Durante a interação, eu consigo acompanhar a movimentação de cada participante no ambiente compartilhado.	① ② ③ ④
Q27. O ambiente colaborativo me permite identificar se o sistema permite diferentes modalidades/dispositivos de acesso (local/remoto, parado/movimento, etc.).	① ② ③ ④
Q28. Ao utilizar o ambiente colaborativo, eu consigo identificar se existe uma dependência entre a aplicação e o local de utilização (recursos/funcionalidades disponíveis depende do local de acesso).	① ② ③ ④
Q29. Ao utilizar o ambiente colaborativo, eu consigo identificar os meios para pré-visualizar (espiar) o conteúdo das tarefas/atividades, sem acessá-los diretamente.	① ② ③ ④
Q30. Ao utilizar o ambiente colaborativo, eu consigo identificar os meios para pré-visualizar (espiar) as tarefas/atividades realizadas pelos demais participantes.	① ② ③ ④

Caderno 4 = {01, 05, 09, 10, 14, 16, 17, 18, 22, 24, 26, 30, 32, 35, 37, 39, 40, 41, 42, 43, 46, 51, 56, 59, 62, 65, 67, 69, 71, 74}

Figure 73 – Awareness Assessment Model Questionnaire: book 4

PARTE 2 – INVENTÁRIO DA PERCEPÇÃO	
Julgue os seguintes itens, considerando a escala: ① Discordo totalmente; ② Discordo; ③ Concordo; e ④ Concordo totalmente.	
Q1. Quando estou interagindo, eu consigo identificar quais são os artefatos ou objetos que estão sendo alterados.	① ② ③ ④
Q2. Ao utilizar o ambiente colaborativo, eu consigo acessar o conteúdo atual dos recursos compartilhados.	① ② ③ ④
Q3. Quando estou interagindo, eu consigo acompanhar o progresso da tarefa/atividade realizada em grupo.	① ② ③ ④
Q4. Ao terminar uma tarefa/atividade conjunta, eu consigo avaliar os resultados obtidos.	① ② ③ ④
Q5. Ao utilizar o ambiente colaborativo, eu consigo identificar os locais onde posso interagir (onde posso realizar as tarefas/atividades).	① ② ③ ④
Q6. Ao utilizar o ambiente colaborativo, eu consigo identificar os recursos/funcionalidades disponíveis para colaborar.	① ② ③ ④
Q7. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem restrições/elementos críticos para a realização das tarefas/atividades.	① ② ③ ④
Q8. Ao utilizar o ambiente colaborativo, eu consigo identificar o relacionamento entre os objetos/recursos do ambiente.	① ② ③ ④
Q9. Durante a interação, eu consigo compreender quais são as próximas tarefas/atividades que devem ser realizadas ou situações futuras que podem ocorrer.	① ② ③ ④
Q10. Quando estou realizando uma atividade em grupo, eu consigo identificar se os demais estão acompanhando as ações realizadas.	① ② ③ ④
Q11. Quando estou realizando uma atividade em grupo, eu consigo acompanhar as alterações realizadas pelos demais participantes.	① ② ③ ④
Q12. Durante a interação, eu consigo identificar como os demais participantes estão controlando suas ações/decisões.	① ② ③ ④
Q13. Ao utilizar o ambiente colaborativo, eu consigo identificar a presença de restrição de acesso (privilégios) do ambiente compartilhado (seja nas tarefas/atividades, objetos/artefatos, ou recursos).	① ② ③ ④
Q14. Ao utilizar o ambiente colaborativo, eu consigo visualizar a identidade dos participantes (quem são?).	① ② ③ ④
Q15. Quando estou utilizando o ambiente colaborativo, eu consigo identificar quais informações estão sendo compartilhadas.	① ② ③ ④
Q16. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem regras distintas (e quais são) para cada participante.	① ② ③ ④
Q17. Durante a interação, eu consigo identificar o que cada participante pode fazer, ver, ou ainda, controlar.	① ② ③ ④
Q18. Durante a realização das tarefas/atividades em grupo, eu consigo identificar o nível de atividade/engajamento de cada participante.	① ② ③ ④
Q19. Ao utilizar o ambiente colaborativo, eu consigo acessar a conectividade de rede.	① ② ③ ④
Q20. Durante a interação, sou notificado se ocorrem atrasos no envio/recebimento das mensagens.	① ② ③ ④
Q21. Durante a interação, eu consigo identificar quais são as expectativas envolvendo cada participante.	① ② ③ ④
Q22. Durante a interação, eu consigo identificar a disponibilidade de informações não verbais para a comunicação (emojis, gestos, etc.).	① ② ③ ④
Q23. Ao utilizar o ambiente colaborativo, eu consigo visualizar a distância de cada participante em relação aos demais.	① ② ③ ④
Q24. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem restrições de espaço envolvidas (e quais são).	① ② ③ ④
Q25. Ao utilizar o ambiente colaborativo, eu consigo identificar como o ambiente está organizado/configurado (topologia).	① ② ③ ④
Q26. Ao utilizar o ambiente colaborativo, eu consigo identificar os atributos dos objetos/recursos ou condições do ambiente.	① ② ③ ④
Q27. O ambiente colaborativo me permite identificar a orientação/direção de cada participante.	① ② ③ ④
Q28. Durante a interação, eu consigo identificar o nível de atenção necessário para a realização da tarefa/atividade.	① ② ③ ④
Q29. O ambiente colaborativo possibilita certa mobilidade ao usuário (acesso por diferentes dispositivos).	① ② ③ ④
Q30. Ao utilizar o ambiente colaborativo, eu consigo identificar onde estão os objetos/artefatos ou recursos no ambiente compartilhado.	① ② ③ ④

Caderno 5 = {02, 03, 06, 08, 12, 19, 20, 21, 23, 27, 28, 29, 31, 33, 34, 36, 38, 44, 47, 49, 53, 55, 57, 58, 60, 61, 64, 66, 68, 75}

Figure 74 – Awareness Assessment Model Questionnaire: book 5

PARTE 2 – INVENTÁRIO DA PERCEPÇÃO	
Julgue os seguintes itens, considerando a escala: ① Discordo totalmente; ② Discordo; ③ Concordo; e ④ Concordo totalmente.	
Q1. Ao utilizar o ambiente colaborativo, eu consigo acessar o conteúdo atual dos recursos compartilhados.	① ② ③ ④
Q2. Quando estou interagindo, eu consigo identificar a motivação para as tarefas/atividades realizadas.	① ② ③ ④
Q3. Quando estou realizando uma atividade em grupo, eu consigo identificar como posso auxiliar os demais participantes.	① ② ③ ④
Q4. Durante a interação, eu consigo identificar o que está acontecendo no ambiente e o que já foi realizado.	① ② ③ ④
Q5. Ao utilizar o ambiente colaborativo, eu consigo identificar os locais onde posso interagir (onde posso realizar as tarefas/atividades).	① ② ③ ④
Q6. Durante a interação, eu consigo identificar se existem outras tarefas/atividades que estão relacionadas ao cenário atual.	① ② ③ ④
Q7. Durante a interação, eu consigo identificar se a tarefa/atividade está sendo realizada de forma coordenada e quem está a coordenado.	① ② ③ ④
Q8. Ao utilizar o ambiente colaborativo, eu consigo identificar os recursos/funcionalidades disponíveis para colaborar.	① ② ③ ④
Q9. Ao utilizar o ambiente colaborativo, eu consigo identificar o relacionamento entre os objetos/recursos do ambiente.	① ② ③ ④
Q10. Durante a interação, eu consigo compreender quais são as próximas tarefas/atividades que devem ser realizadas ou situações futuras que podem ocorrer.	① ② ③ ④
Q11. Ao utilizar o ambiente colaborativo, eu consigo identificar a resposta das ações realizadas por mim (feedback).	① ② ③ ④
Q12. Quando estou realizando uma atividade em grupo, eu consigo identificar se os demais estão acompanhando as ações realizadas.	① ② ③ ④
Q13. Durante a interação, eu consigo identificar como os demais participantes estão controlando suas ações/decisões.	① ② ③ ④
Q14. Ao utilizar o ambiente colaborativo, eu consigo identificar a presença de restrição de acesso (privilégios) do ambiente compartilhado (seja nas tarefas/atividades, objetos/artefatos, ou recursos).	① ② ③ ④
Q15. Durante a realização das tarefas/atividades em grupo, eu consigo identificar o nível de atividade/engajamento de cada participante.	① ② ③ ④
Q16. Ao utilizar o ambiente colaborativo, eu consigo identificar o estado/situação atual de cada participante.	① ② ③ ④
Q17. Ao utilizar o ambiente colaborativo, eu consigo acessar a conectividade de rede.	① ② ③ ④
Q18. Durante a interação, eu consigo visualizar quando as mensagens são enviadas/recebidas pelos demais participantes.	① ② ③ ④
Q19. Ao utilizar o ambiente colaborativo, eu consigo identificar os meios disponíveis para conectar e interagir com os demais participantes.	① ② ③ ④
Q20. O ambiente colaborativo me permite identificar os meios disponíveis para conversar com os demais participantes (estabelecer uma comunicação).	① ② ③ ④
Q21. Durante a interação, eu consigo identificar o estado emocional de cada participante.	① ② ③ ④
Q22. Durante a interação, eu consigo identificar a disponibilidade de informações não verbais para a comunicação (emojis, gestos, etc.).	① ② ③ ④
Q23. Ao utilizar o ambiente colaborativo, eu consigo visualizar a distância de cada participante em relação aos demais.	① ② ③ ④
Q24. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem restrições de espaço envolvidas (e quais são).	① ② ③ ④
Q25. Ao utilizar o ambiente colaborativo, eu consigo identificar os atributos dos objetos/recursos ou condições do ambiente.	① ② ③ ④
Q26. O ambiente colaborativo me permite identificar o alcance de cada participante (onde pode ir, o que pode acessar).	① ② ③ ④
Q27. Quando estou interagindo, eu consigo identificar quem está falando com quem (comunicação verbal).	① ② ③ ④
Q28. Durante a interação, eu consigo identificar onde cada participante está olhando.	① ② ③ ④
Q29. O ambiente colaborativo me permite visualizar o ambiente compartilhado de forma simplificada (miniatura, mapa, ou outro similar).	① ② ③ ④
Q30. Ao utilizar o ambiente colaborativo, eu consigo identificar onde estão os objetos/artefatos ou recursos no ambiente compartilhado.	① ② ③ ④

Caderno 6 = {03, 04, 07, 11, 12, 13, 15, 19, 21, 23, 25, 27, 29, 31, 44, 45, 47, 48, 50, 52, 54, 55, 57, 58, 61, 63, 70, 72, 73, 75}

Figure 75 – Awareness Assessment Model Questionnaire: book 6

PARTE 2 – INVENTÁRIO DA PERCEPÇÃO	
Julgue os seguintes itens, considerando a escala: ① Discordo totalmente; ② Discordo; ③ Concordo; e ④ Concordo totalmente.	
Q1. Quando estou interagindo, eu consigo identificar quais são os artefatos ou objetos que estão sendo alterados.	① ② ③ ④
Q2. Quando estou interagindo, eu consigo identificar a motivação para as tarefas/atividades realizadas.	① ② ③ ④
Q3. Quando estou interagindo, eu consigo acompanhar o progresso da tarefa/atividade realizada em grupo.	① ② ③ ④
Q4. Quando estou realizando uma atividade em grupo, eu consigo identificar como posso auxiliar os demais participantes.	① ② ③ ④
Q5. Ao terminar uma tarefa/atividade conjunta, eu consigo avaliar os resultados obtidos.	① ② ③ ④
Q6. Durante a interação, eu consigo identificar o que está acontecendo no ambiente e o que já foi realizado.	① ② ③ ④
Q7. Durante a interação, eu consigo identificar se existem outras tarefas/atividades que estão relacionadas ao cenário atual.	① ② ③ ④
Q8. Durante a interação, eu consigo identificar se a tarefa/atividade está sendo realizada de forma coordenada e quem está a coordenado.	① ② ③ ④
Q9. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem restrições/elementos críticos para a realização das tarefas/atividades.	① ② ③ ④
Q10. Ao utilizar o ambiente colaborativo, eu consigo identificar a resposta das ações realizadas por mim (feedback).	① ② ③ ④
Q11. Quando estou realizando uma atividade em grupo, eu consigo acompanhar as alterações realizadas pelos demais participantes.	① ② ③ ④
Q12. Ao utilizar o ambiente colaborativo, eu consigo visualizar a identidade dos participantes (quem são?).	① ② ③ ④
Q13. Quando estou utilizando o ambiente colaborativo, eu consigo identificar quais informações estão sendo compartilhadas.	① ② ③ ④
Q14. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem regras distintas (e quais são) para cada participante.	① ② ③ ④
Q15. Durante a interação, eu consigo identificar o que cada participante pode fazer, ver, ou ainda, controlar.	① ② ③ ④
Q16. Ao utilizar o ambiente colaborativo, eu consigo identificar o estado/situação atual de cada participante.	① ② ③ ④
Q17. Durante a interação, eu consigo visualizar quando as mensagens são enviadas/recebidas pelos demais participantes.	① ② ③ ④
Q18. Durante a interação, sou notificado se ocorrem atrasos no envio/recebimento das mensagens.	① ② ③ ④
Q19. Ao utilizar o ambiente colaborativo, eu consigo identificar os meios disponíveis para conectar e interagir com os demais participantes.	① ② ③ ④
Q20. O ambiente colaborativo me permite identificar os meios disponíveis para conversar com os demais participantes (estabelecer uma comunicação).	① ② ③ ④
Q21. Durante a interação, eu consigo identificar quais são as expectativas envolvendo cada participante.	① ② ③ ④
Q22. Durante a interação, eu consigo identificar o estado emocional de cada participante.	① ② ③ ④
Q23. Ao utilizar o ambiente colaborativo, eu consigo identificar como o ambiente está organizado/configurado (topologia).	① ② ③ ④
Q24. O ambiente colaborativo me permite identificar o alcance de cada participante (onde pode ir, o que pode acessar).	① ② ③ ④
Q25. O ambiente colaborativo me permite identificar a orientação/direção de cada participante.	① ② ③ ④
Q26. Durante a interação, eu consigo identificar o nível de atenção necessário para a realização da tarefa/atividade.	① ② ③ ④
Q27. O ambiente colaborativo possibilita certa mobilidade ao usuário (acesso por diferentes dispositivos).	① ② ③ ④
Q28. Quando estou interagindo, eu consigo identificar quem está falando com quem (comunicação verbal).	① ② ③ ④
Q29. Durante a interação, eu consigo identificar onde cada participante está olhando.	① ② ③ ④
Q30. O ambiente colaborativo me permite visualizar o ambiente compartilhado de forma simplificada (miniatura, mapa, ou outro similar).	① ② ③ ④

Caderno 7 = {02, 04, 06, 07, 08, 11, 13, 15, 20, 25, 28, 33, 34, 36, 38, 45, 48, 49, 50, 52, 53, 54, 60, 63, 64, 66, 68, 70, 72, 73}

Figure 76 – Awareness Assessment Model Questionnaire: book 7

PARTE 2 – INVENTÁRIO DA PERCEPÇÃO	
Julgue os seguintes itens, considerando a escala: ① Discordo totalmente; ② Discordo; ③ Concordo; e ④ Concordo totalmente.	
Q1. Quando estou interagindo, eu consigo identificar a motivação para as tarefas/atividades realizadas.	① ② ③ ④
Q2. Ao utilizar o ambiente colaborativo, eu consigo identificar o tempo necessário/disponível para realizar a tarefa/atividade.	① ② ③ ④
Q3. Quando estou realizando uma atividade em grupo, eu consigo identificar como posso auxiliar os demais participantes.	① ② ③ ④
Q4. Quando estou realizando uma atividade em grupo, eu consigo compreender quais são os passos/ações que devem ser realizadas.	① ② ③ ④
Q5. Durante a interação, eu consigo identificar o que está acontecendo no ambiente e o que já foi realizado.	① ② ③ ④
Q6. Durante a interação, eu consigo identificar se existem outras tarefas/atividades que estão relacionadas ao cenário atual.	① ② ③ ④
Q7. Durante a interação, eu consigo identificar se a tarefa/atividade está sendo realizada de forma coordenada e quem está a coordenado.	① ② ③ ④
Q8. Durante a interação, eu consigo identificar como a tarefa/atividade atual está relacionada ao cenário atual (tarefas/atividades conjuntas).	① ② ③ ④
Q9. Ao utilizar o ambiente colaborativo, eu consigo identificar a presença de artefatos/objetos necessários para colaborar.	① ② ③ ④
Q10. Durante a interação, eu consigo compreender o significado das ações realizadas (o que está acontecendo no ambiente).	① ② ③ ④
Q11. Ao utilizar o ambiente colaborativo, eu consigo identificar a resposta das ações realizadas por mim (feedback).	① ② ③ ④
Q12. Ao utilizar o ambiente colaborativo, eu consigo identificar a resposta das ações realizadas pelos demais participantes (feedthrought).	① ② ③ ④
Q13. O ambiente colaborativo me permite identificar se existem mecanismos de controle de acesso aos recursos/funcionalidades e como posso acessá-los.	① ② ③ ④
Q14. O ambiente colaborativo me permite identificar as preferências individuais de cada participante.	① ② ③ ④
Q15. Durante a interação, eu consigo identificar as responsabilidades que cada participante pode assumir.	① ② ③ ④
Q16. Quando estou interagindo, eu consigo identificar quais são as influências/decisões de cada participante.	① ② ③ ④
Q17. Durante a interação, eu consigo expressar minhas intenções (o que pretendo fazer) no ambiente e consigo identificar as intenções dos demais participantes.	① ② ③ ④
Q18. Ao utilizar o ambiente colaborativo, eu consigo identificar o estado/situação atual de cada participante.	① ② ③ ④
Q19. Durante a interação, eu consigo visualizar quando as mensagens são enviadas/recebidas pelos demais participantes.	① ② ③ ④
Q20. Ao utilizar o ambiente colaborativo, eu consigo identificar os meios disponíveis para conectar e interagir com os demais participantes.	① ② ③ ④
Q21. O ambiente colaborativo me permite identificar os meios disponíveis para conversar com os demais participantes (estabelecer uma comunicação).	① ② ③ ④
Q22. Durante a interação, eu consigo identificar o estado emocional de cada participante.	① ② ③ ④
Q23. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem diferentes locais para colaboração (e quais são).	① ② ③ ④
Q24. O ambiente colaborativo me permite identificar o que cada participante pode ver.	① ② ③ ④
Q25. O ambiente colaborativo me permite identificar o alcance de cada participante (onde pode ir, o que pode acessar).	① ② ③ ④
Q26. Quando estou interagindo, eu consigo identificar quem está falando com quem (comunicação verbal).	① ② ③ ④
Q27. Ao utilizar o ambiente colaborativo, eu consigo identificar os meios para pré-visualizar (espiar) o conteúdo das tarefas/atividades, sem acessá-los diretamente.	① ② ③ ④
Q28. Durante a interação, eu consigo identificar onde cada participante está olhando.	① ② ③ ④
Q29. O ambiente colaborativo me permite visualizar o ambiente compartilhado de forma simplificada (miniatura, mapa, ou outro similar).	① ② ③ ④
Q30. Ao utilizar o ambiente colaborativo, eu consigo identificar os meios para pré-visualizar (espiar) as tarefas/atividades realizadas pelos demais participantes.	① ② ③ ④

Caderno 8 = {04, 05, 07, 10, 11, 13, 15, 16, 18, 22, 25, 26, 32, 35, 37, 40, 41, 45, 48, 50, 52, 54, 59, 62, 63, 70, 71, 72, 73, 74}

Figure 77 – Awareness Assessment Model Questionnaire: book 8

PARTE 2 – INVENTÁRIO DA PERCEPÇÃO	
Julgue os seguintes itens, considerando a escala: ① Discordo totalmente; ② Discordo; ③ Concordo; e ④ Concordo totalmente.	
Q1. O ambiente colaborativo me permite identificar o objetivo das tarefas/atividades realizadas.	① ② ③ ④
Q2. Ao utilizar o ambiente colaborativo, eu consigo acessar o conteúdo atual dos recursos compartilhados.	① ② ③ ④
Q3. O ambiente colaborativo me permite identificar quem está conduzindo as tarefas/atividades (autoria).	① ② ③ ④
Q4. Ao utilizar o ambiente colaborativo, eu consigo identificar os locais onde posso interagir (onde posso realizar as tarefas/atividades).	① ② ③ ④
Q5. Durante a interação, eu consigo identificar se os outros participantes estão engajados com a tarefa/atividade atual ou se estão realizando outras tarefas/atividades em paralelo.	① ② ③ ④
Q6. Ao utilizar o ambiente colaborativo, eu consigo identificar as ferramentas e materiais disponíveis para colaborar.	① ② ③ ④
Q7. Ao utilizar o ambiente colaborativo, eu consigo identificar os recursos/funcionalidades disponíveis para colaborar.	① ② ③ ④
Q8. Ao utilizar o ambiente colaborativo, eu consigo identificar o relacionamento entre os objetos/recursos do ambiente.	① ② ③ ④
Q9. Durante a interação, eu consigo compreender quais são as próximas tarefas/atividades que devem ser realizadas ou situações futuras que podem ocorrer.	① ② ③ ④
Q10. Quando estou realizando uma atividade em grupo, eu consigo identificar a compreensão dos demais participantes envolvidos.	① ② ③ ④
Q11. Quando estou realizando uma atividade em grupo, eu consigo identificar se os demais estão acompanhando as ações realizadas.	① ② ③ ④
Q12. Durante a interação, eu consigo identificar como os demais participantes estão controlando suas ações/decisões.	① ② ③ ④
Q13. Ao utilizar o ambiente colaborativo, eu consigo identificar a presença de restrição de acesso (privilégios) do ambiente compartilhado (seja nas tarefas/atividades, objetos/artefatos, ou recursos).	① ② ③ ④
Q14. Quando estou interagindo, eu consigo identificar o que eu sei sobre a tarefa/atividade atual e como eu posso ajudar.	① ② ③ ④
Q15. Ao utilizar o ambiente colaborativo, eu consigo identificar a disponibilidade de cada participante.	① ② ③ ④
Q16. Ao utilizar o ambiente colaborativo, eu consigo identificar a presença de cada participante no ambiente.	① ② ③ ④
Q17. Durante a realização das tarefas/atividades em grupo, eu consigo identificar o nível de atividade/engajamento de cada participante.	① ② ③ ④
Q18. Ao utilizar o ambiente colaborativo, eu consigo identificar o modo de trabalho (síncrono ou assíncrono).	① ② ③ ④
Q19. Ao utilizar o ambiente colaborativo, eu consigo acessar a conectividade de rede.	① ② ③ ④
Q20. Durante a interação, eu consigo identificar quem está conversando, trocando ideias, ou de quem é a vez de falar.	① ② ③ ④
Q21. Durante a interação, eu consigo identificar a disponibilidade de informações não verbais para a comunicação (emojis, gestos, etc.).	① ② ③ ④
Q22. Ao utilizar o ambiente colaborativo, eu consigo visualizar a localização física/virtual dos demais participantes.	① ② ③ ④
Q23. Ao utilizar o ambiente colaborativo, eu consigo visualizar a distância de cada participante em relação aos demais.	① ② ③ ④
Q24. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem restrições de espaço envolvidas (e quais são).	① ② ③ ④
Q25. Ao utilizar o ambiente colaborativo, eu consigo identificar os atributos dos objetos/recursos ou condições do ambiente.	① ② ③ ④
Q26. Durante a interação, eu consigo acompanhar a movimentação de cada participante no ambiente compartilhado.	① ② ③ ④
Q27. Durante a interação, eu consigo identificar o nível de atenção necessário para a realização da tarefa/atividade.	① ② ③ ④
Q28. O ambiente colaborativo me permite identificar se o sistema permite diferentes modalidades/dispositivos de acesso (local/remoto, parado/movimento, etc.).	① ② ③ ④
Q29. Ao utilizar o ambiente colaborativo, eu consigo identificar se existe uma dependência entre a aplicação e o local de utilização (recursos/funcionalidades disponíveis depende do local de acesso).	① ② ③ ④
Q30. Ao utilizar o ambiente colaborativo, eu consigo identificar onde estão os objetos/artefatos ou recursos no ambiente compartilhado.	① ② ③ ④

Caderno 9 = {01, 03, 09, 12, 14, 17, 19, 21, 23, 24, 27, 29, 30, 31, 39, 42, 43, 44, 46, 47, 51, 55, 56, 57, 58, 61, 65, 67, 69, 75}

Figure 78 – Awareness Assessment Model Questionnaire: book 9

PARTE 2 – INVENTÁRIO DA PERCEPÇÃO	
Julgue os seguintes itens, considerando a escala: ① Discordo totalmente; ② Discordo; ③ Concordo; e ④ Concordo totalmente.	
Q1. Quando estou interagindo, eu consigo identificar quais são os artefatos ou objetos que estão sendo alterados.	① ② ③ ④
Q2. Ao utilizar o ambiente colaborativo, eu consigo identificar o tempo necessário/disponível para realizar a tarefa/atividade.	① ② ③ ④
Q3. Quando estou interagindo, eu consigo acompanhar o progresso das tarefa/atividade realizadas em grupo.	① ② ③ ④
Q4. Ao terminar uma tarefa/atividade conjunta, eu consigo avaliar os resultados obtidos.	① ② ③ ④
Q5. Quando estou realizando uma atividade em grupo, eu consigo compreender quais são os passos/ações que devem ser realizadas.	① ② ③ ④
Q6. Durante a interação, eu consigo identificar como a tarefa/atividade atual está relacionada ao cenário atual (tarefas/atividades conjuntas).	① ② ③ ④
Q7. Ao utilizar o ambiente colaborativo, eu consigo identificar a presença de artefatos/objetos necessários para colaborar.	① ② ③ ④
Q8. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem restrições/elementos críticos para a realização das tarefas/atividades.	① ② ③ ④
Q9. Durante a interação, eu consigo compreender o significado das ações realizadas (o que está acontecendo no ambiente).	① ② ③ ④
Q10. Ao utilizar o ambiente colaborativo, eu consigo identificar a resposta das ações realizadas pelos demais participantes (feedthrought).	① ② ③ ④
Q11. Quando estou realizando uma atividade em grupo, eu consigo acompanhar as alterações realizadas pelos demais participantes.	① ② ③ ④
Q12. O ambiente colaborativo me permite identificar se existem mecanismos de controle de acesso aos recursos/funcionalidades e como posso acessá-los.	① ② ③ ④
Q13. Ao utilizar o ambiente colaborativo, eu consigo visualizar a identidade dos participantes (quem são?).	① ② ③ ④
Q14. Quando estou utilizando o ambiente colaborativo, eu consigo identificar quais informações estão sendo compartilhadas.	① ② ③ ④
Q15. O ambiente colaborativo me permite identificar as preferências individuais de cada participante.	① ② ③ ④
Q16. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem regras distintas (e quais são) para cada participante.	① ② ③ ④
Q17. Durante a interação, eu consigo identificar as responsabilidades que cada participante pode assumir.	① ② ③ ④
Q18. Durante a interação, eu consigo identificar o que cada participante pode fazer, ver, ou ainda, controlar.	① ② ③ ④
Q19. Quando estou interagindo, eu consigo identificar quais são as influências/decisões de cada participante.	① ② ③ ④
Q20. Durante a interação, eu consigo expressar minhas intenções (o que pretendo fazer) no ambiente e consigo identificar as intenções dos demais participantes.	① ② ③ ④
Q21. Durante a interação, sou notificado se ocorrem atrasos no envio/recebimento das mensagens.	① ② ③ ④
Q22. Durante a interação, eu consigo identificar quais são as expectativas envolvendo cada participante.	① ② ③ ④
Q23. Ao utilizar o ambiente colaborativo, eu consigo identificar se existem diferentes locais para colaboração (e quais são).	① ② ③ ④
Q24. Ao utilizar o ambiente colaborativo, eu consigo identificar como o ambiente está organizado/configurado (topologia).	① ② ③ ④
Q25. O ambiente colaborativo me permite identificar o que cada participante pode ver.	① ② ③ ④
Q26. O ambiente colaborativo me permite identificar a orientação/direção de cada participante.	① ② ③ ④
Q27. Durante a interação, eu consigo identificar o nível de atenção necessário para a realização da tarefa/atividade.	① ② ③ ④
Q28. O ambiente colaborativo possibilita certa mobilidade ao usuário (acesso por diferentes dispositivos).	① ② ③ ④
Q29. Ao utilizar o ambiente colaborativo, eu consigo identificar os meios para pré-visualizar (espiar) o conteúdo das tarefas/atividades, sem acessá-los diretamente.	① ② ③ ④
Q30. Ao utilizar o ambiente colaborativo, eu consigo identificar os meios para pré-visualizar (espiar) as tarefas/atividades realizadas pelos demais participantes.	① ② ③ ④

Caderno 10 = {02, 05, 06, 08, 10, 16, 18, 20, 22, 26, 28, 32, 33, 34, 35, 36, 37, 38, 40, 41, 49, 53, 59, 60, 62, 64, 66, 68, 71, 74}

Figure 79 – Awareness Assessment Model Questionnaire: book 10

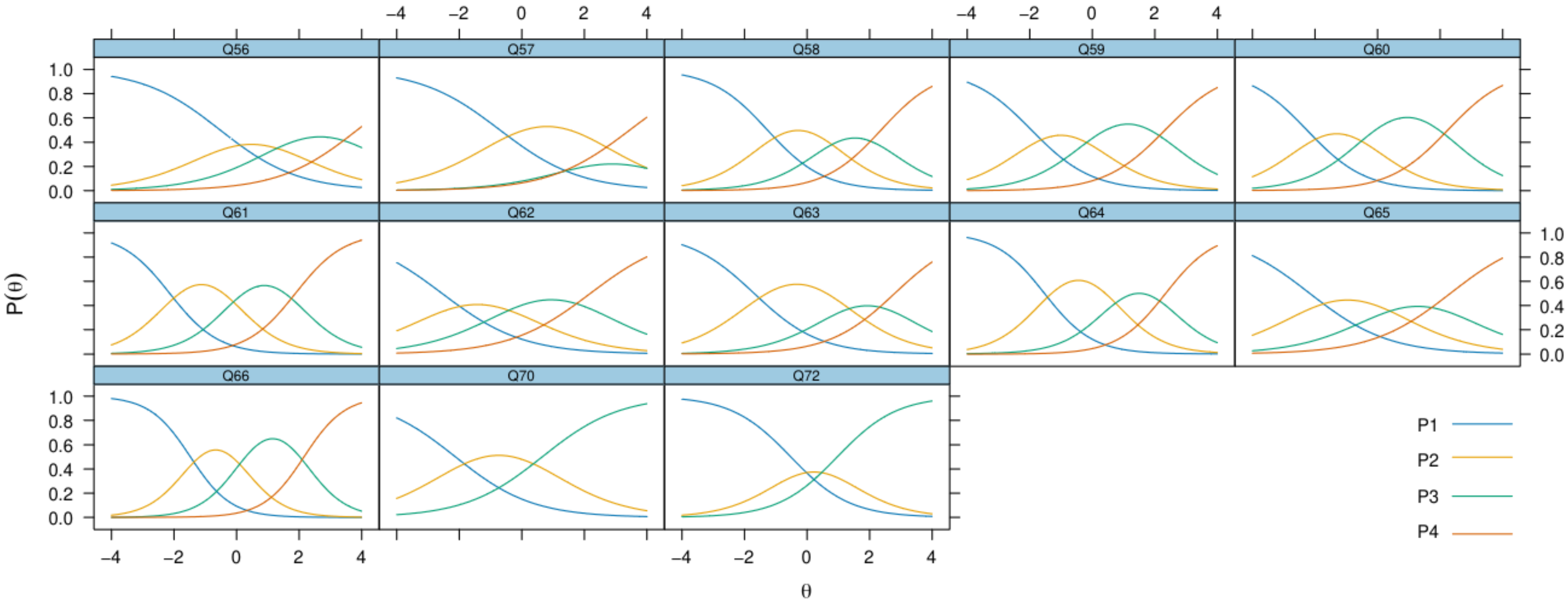


Figure 80 – Contextual Awareness Items' information curves (Moodle scenario)

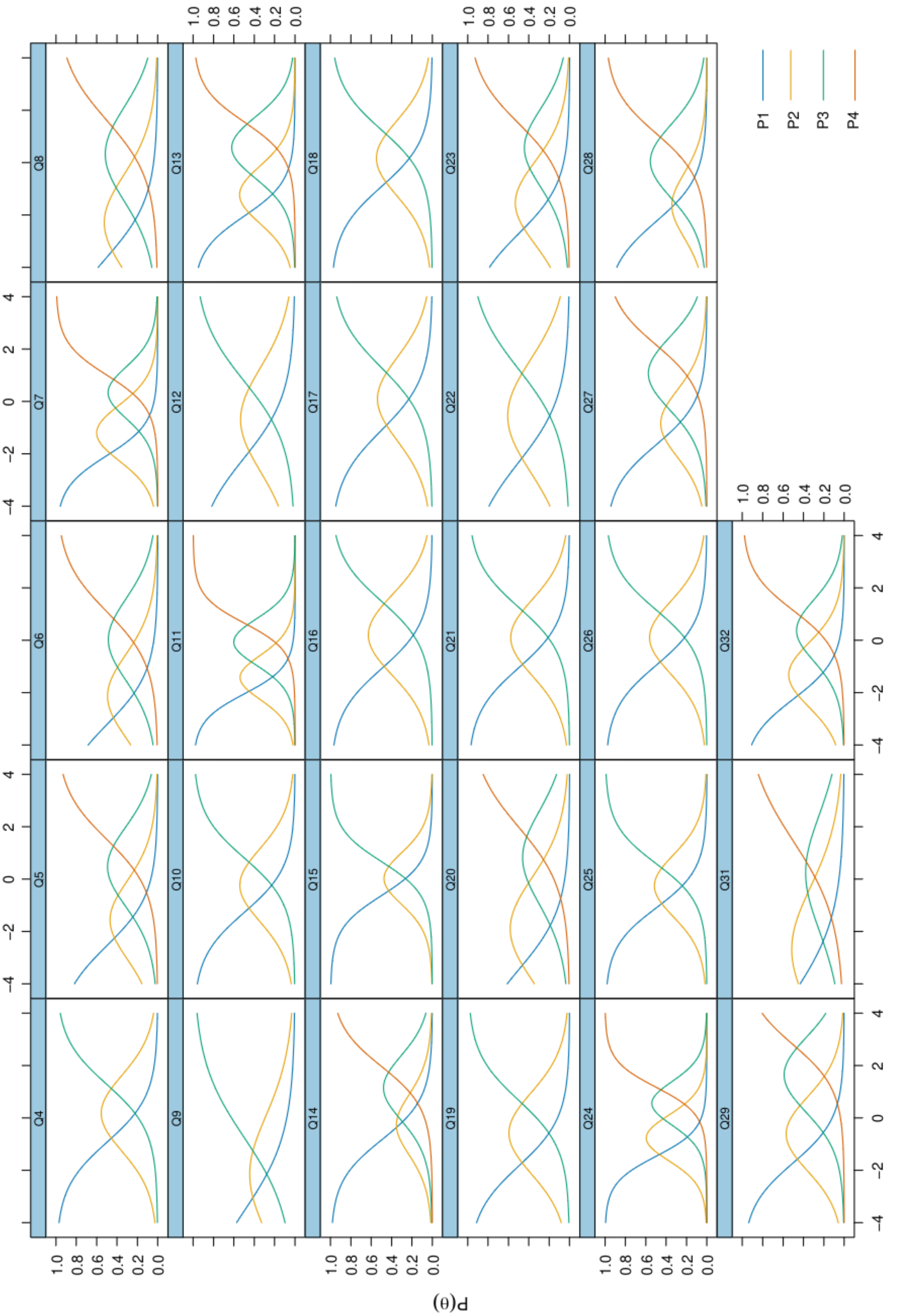


Figure 81 – Workspace Awareness Items' information curves (Moodle scenario)

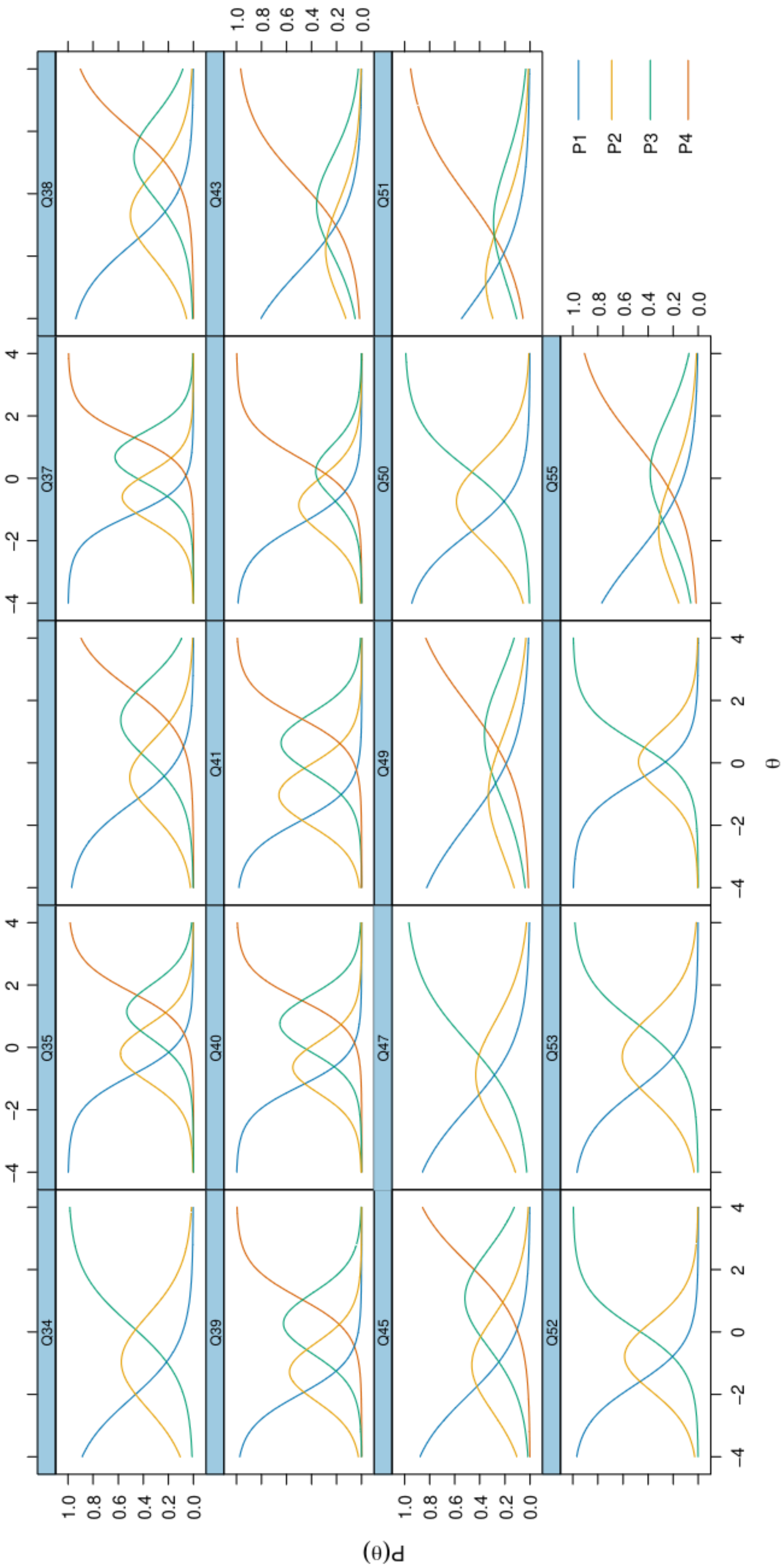


Figure 82 – Collaboration Awareness Items' information curves (Moodle scenario)

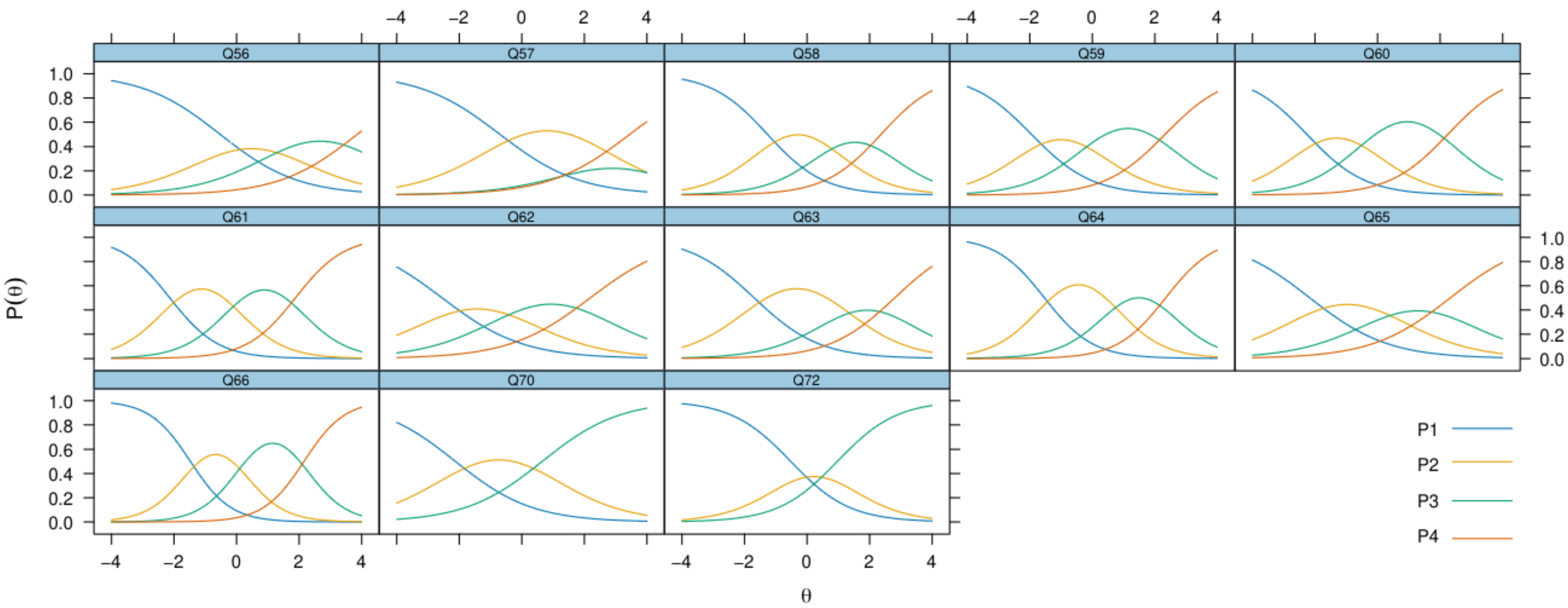


Figure 83 – Contextual Awareness Items’ information curves (videoconferencing scenario)

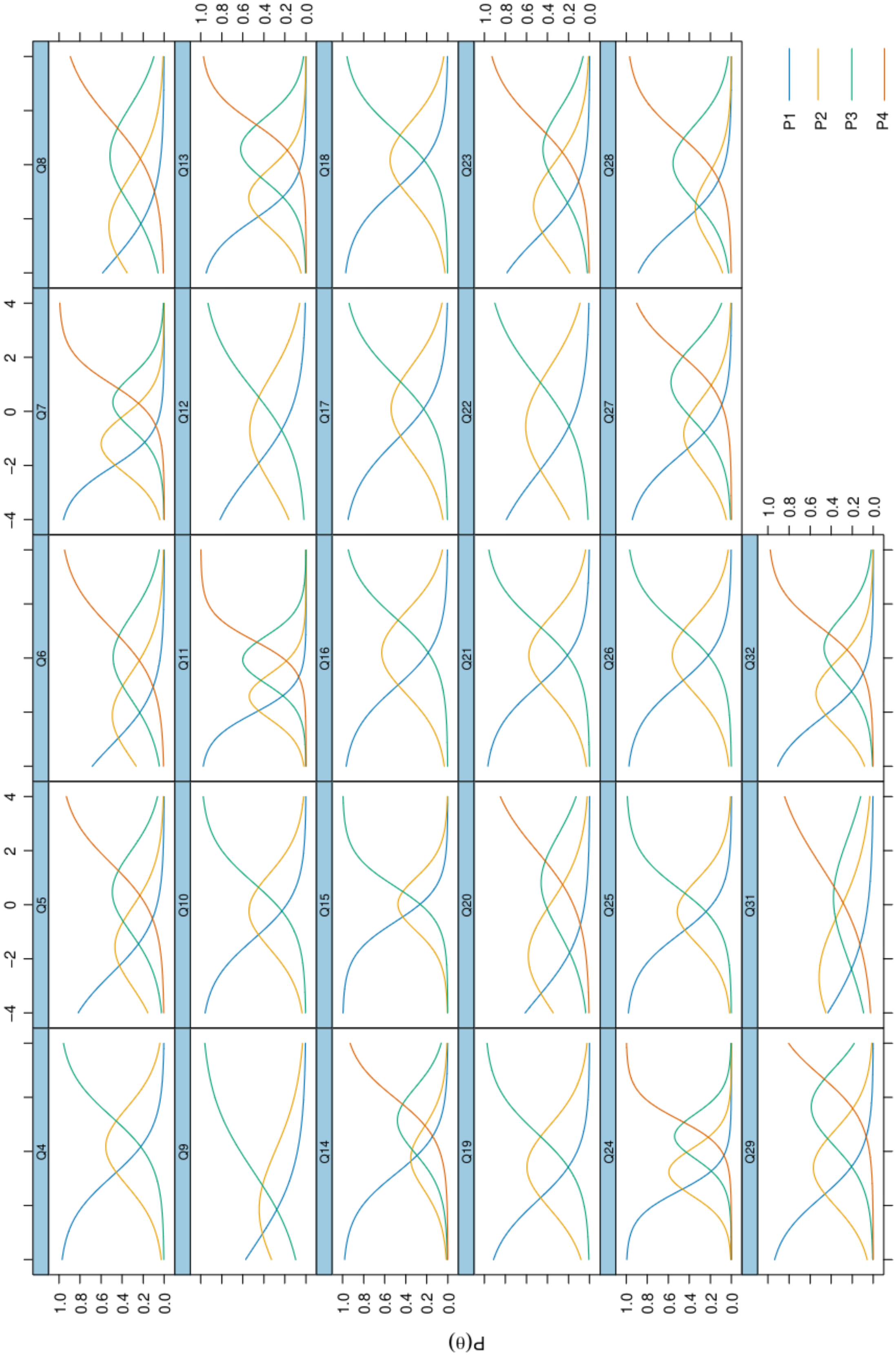


Figure 84 – Workspace Awareness Items' information curves (videoconferencing scenario)

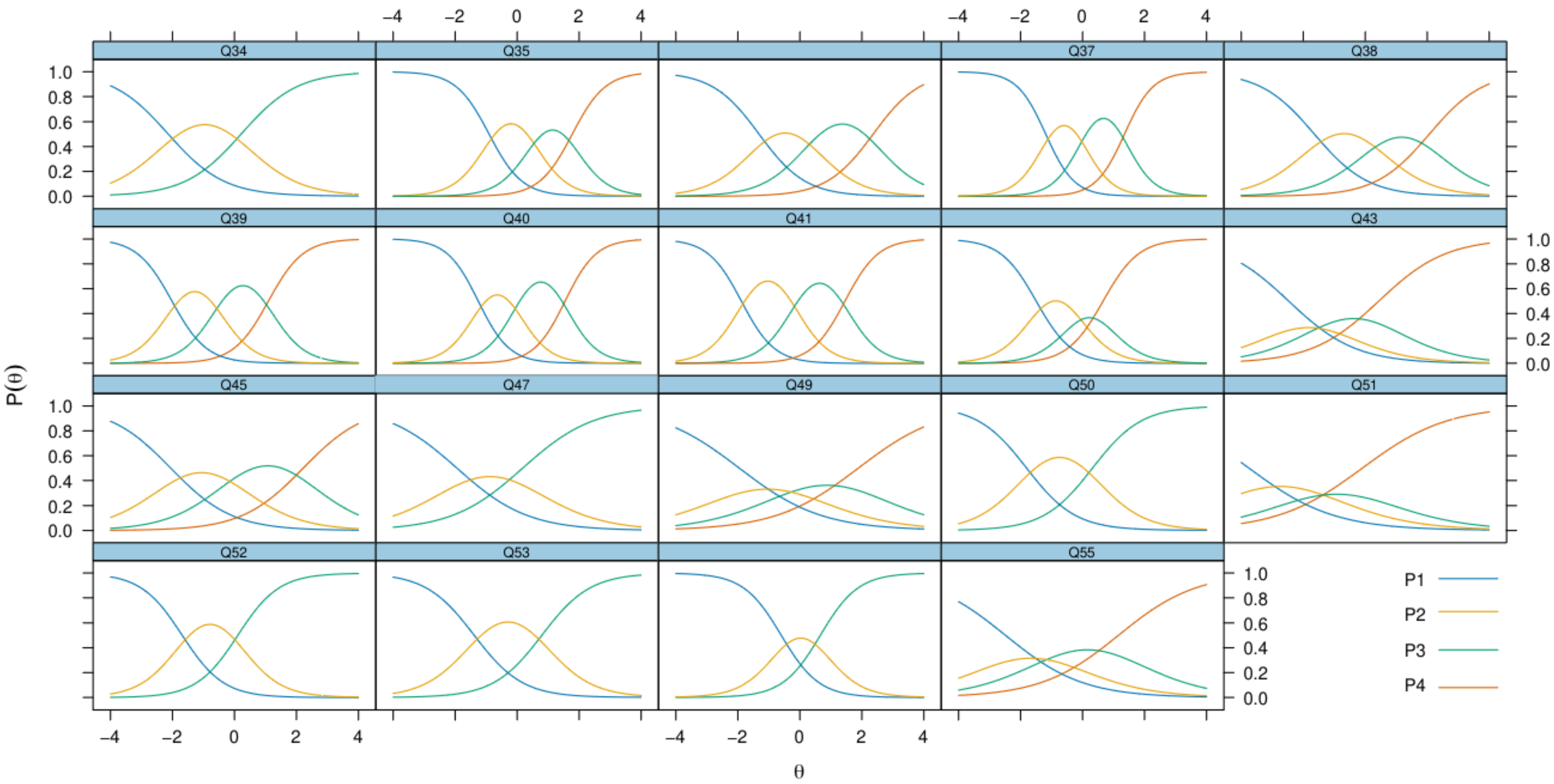


Figure 85 – Collaboration Awareness Items' information curves (videoconferencing scenario)

APPENDIX F – GLOBAL AWARENESS SCALE MATERIALS

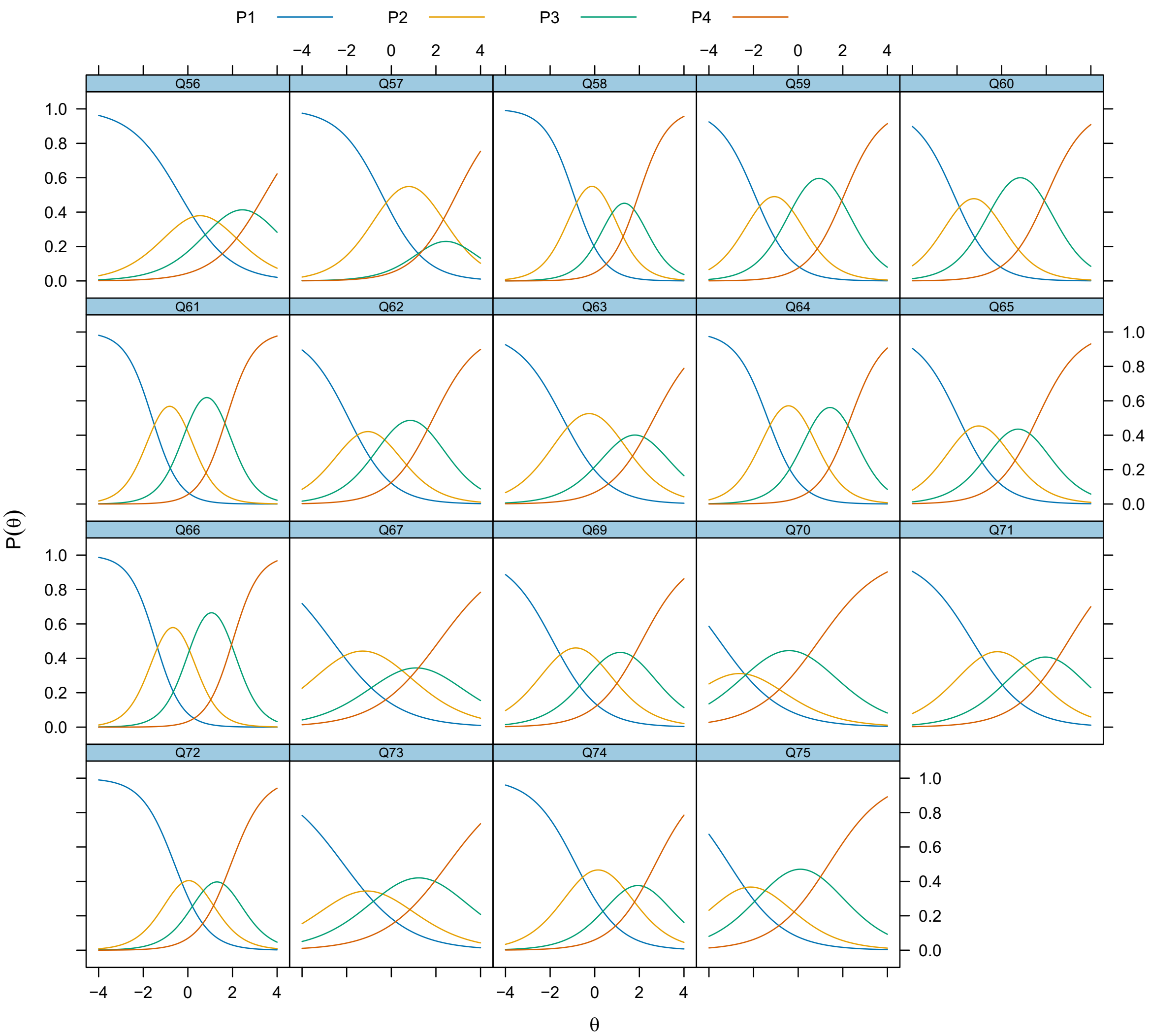


Figure 86 – Contextual Awareness Items' information curves (global scale)

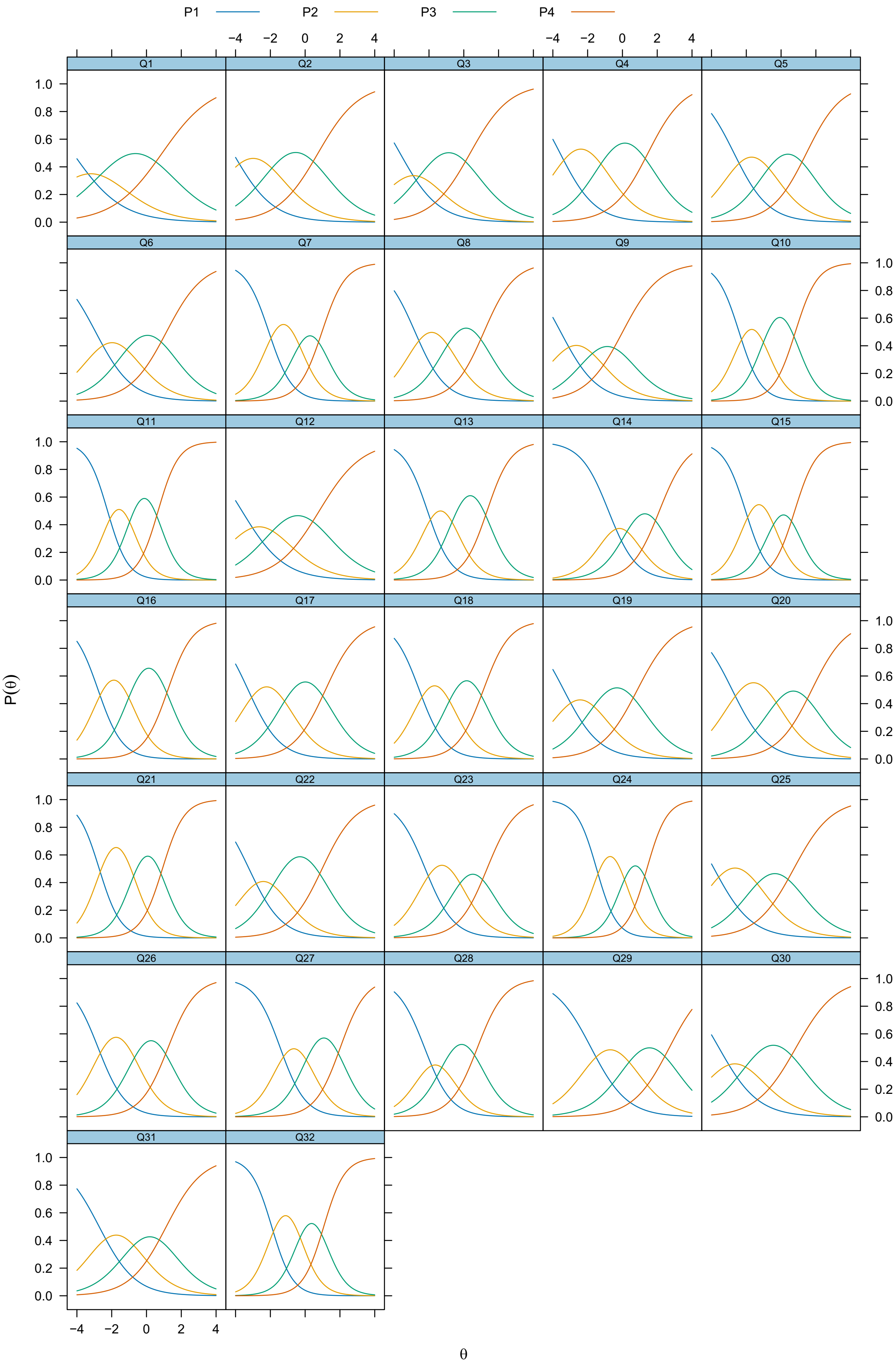


Figure 87 – Workspace Awareness Items' information curves (global scale)

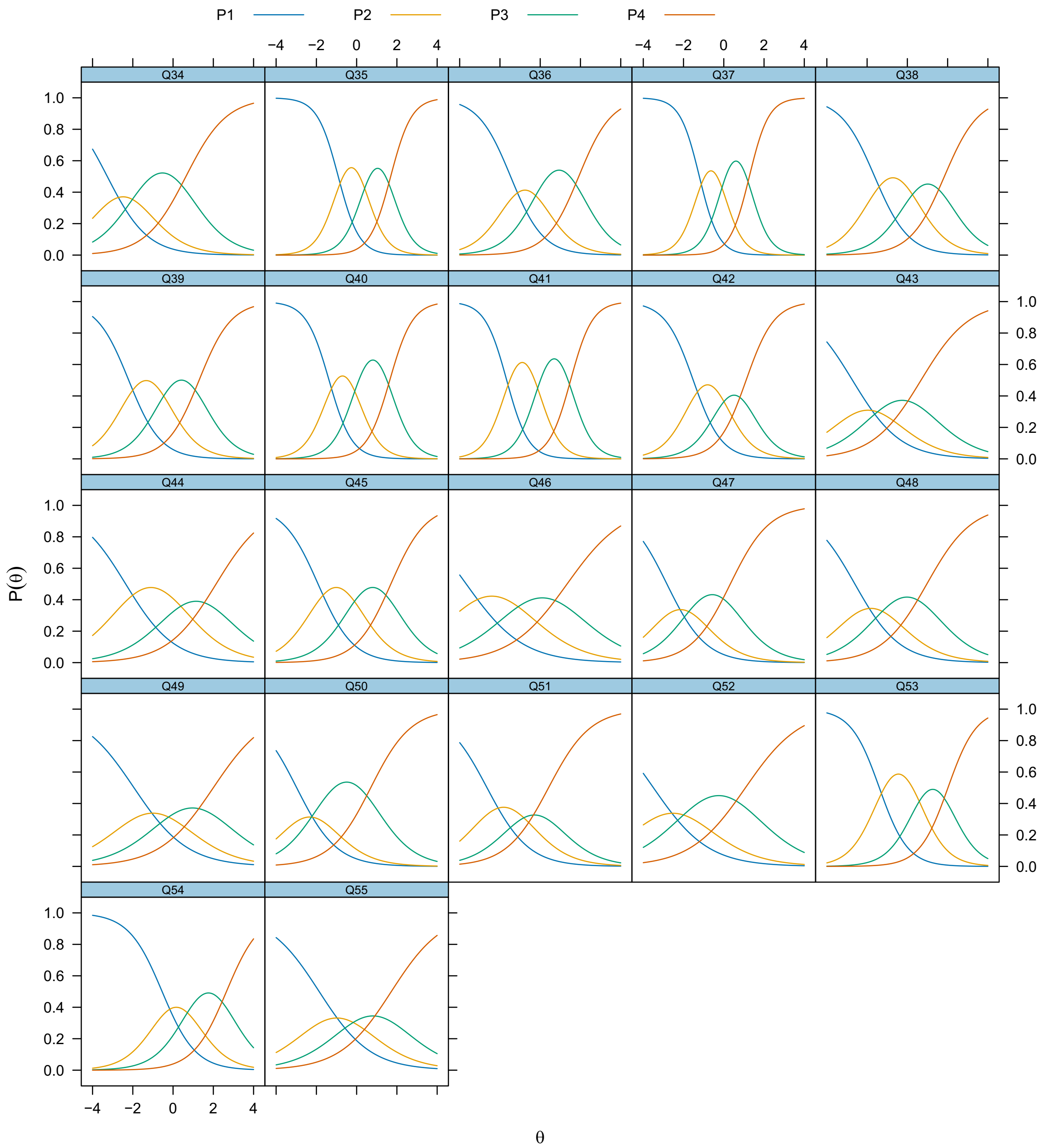


Figure 88 – Collaboration Awareness Items' information curves (global scale)