IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 19 June 2025, accepted 30 July 2025, date of publication 8 August 2025, date of current version 15 August 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3597072

RESEARCH ARTICLE

User-Perceived Quality and Functional Suitability of Representations of DEMO's Process and Fact Models: Conclusions From a Family of Three Studies

DULCE PACHECO[1,2,3], DAVID AVEIRO[1,3,4], DUARTE PINTO[1,3], AND VÍTOR FREITAS[1,3,4]

[1]Enterprise Engineering Laboratory, ARDITI-Regional Agency for the Development of Research, Technology and Innovation, 9020-105 Funchal, Portugal
[2]School of Technology and Management, University of Madeira, 9020-105 Funchal, Portugal
[3]NOVA-LINCS, Universidade NOVA de Lisboa, Campus da Caparica, 2829-516 Costa da Caparica, Portugal
[4]Faculty of Exact Sciences and Engineering, University of Madeira, 9020-105 Funchal, Portugal

Corresponding author: Dulce Pacheco (dulce.pacheco@arditi.pt)

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT Clear and cognitively compelling diagrams representing organizational reality are crucial for structured communication, proper information systems implementation, and ultimately, organizational success. Design and Engineering Methodology for Organizations (DEMO), a standard from the field of Enterprise Engineering, provides representations for Process and Fact Models. However, the complexity of its notations can impede stakeholder understanding. New representations for these models have been proposed, but lacked formal evaluation across diverse user groups. This paper addresses this gap by assessing user perception of both standard and newly proposed DEMO representations. We conducted a family of three evaluation studies using mixed research methods across distinct user groups ($N = 32$). Results show a unanimous preference for the new representations in terms of perceived quality and functional suitability, supporting most of our hypotheses. Our study does not support the claim that familiarity with the modeling language improves perception of empirical quality. This research expands the knowledge base on the underexplored area of user experience within DEMO models and representations.

INDEX TERMS Business process management, business process re-engineering, cognitive processes, data models, information representation, knowledge management, modeling, process modeling, requirements management, user experience.

I. INTRODUCTION

Efficient information and communication systems are vital for organizational success, driving change, evolution, and productivity [1]. However, the effectiveness of these structures hinges on the underlying design of both the organizational and information systems [1]. Flawed or incomplete design can hinder productivity and, consequently, organizational success.

Within enterprises, it is crucial to represent its information system and respective processes and facts clearly. Standards for representing organizational processes and facts are provided by modeling methodologies, such as Design and Engineering Methodology for Organizations (DEMO) [1]. DEMO's interrelated models and diagrams allow us to specify a concise, comprehensive, and technology-independent representation of organizational reality [1]. DEMO's Specification Language specifies the notation to be used in the representation of its models [1], namely for the Process Model to represent business transactions, and the Fact Model to represent facts.

Nevertheless, complex modeling notations can hamper stakeholders' understanding, especially of process and fact models, which are critical for successful system planning

The associate editor coordinating the review of this manuscript and approving it for publication was Claudia Raibulet.

and implementation [2], [3], [4]. Studies highlight the toil of process modeling, mainly owing to the miscellaneous competencies of those involved and the inherent difficulty of understanding the notation semantics [2], [4], [5]. Consequently, guaranteeing model comprehensibility to stakeholders has become critical.

Modeling an organization (or part of it) requires a comprehensive methodology to represent its processes and information. The perceived quality and suitability of representations affect a system's adequacy and implementation effectiveness [2], [6]. Imprecise or ambiguous representations can lead to system inadequacies and jeopardize work effectiveness. Therefore, the cognitive effectiveness of diagrams that represent business processes and facts is particularly important [7], [8].

Within the topic of process modeling quality, most research is focused on empirical and pragmatic quality aspects, specifically with respect to improving the understandability or readability of models [4], [6]. However, scarce studies focused on the stakeholders' perceptions [6]. Also, there is no generally accepted framework or definition of process modeling quality [4], [6].

Business process models should be designed as valuable communication tools for stakeholders [3]. This importance goes beyond following syntactic rules, the goal is to achieve *fitness for use* by the audience [3]. According to Krippendorff [9], design should be defined as the creation of meaning, whereas the value of any design is determined in correspondence with the perceptions/meanings that users attribute to it.

In our applied research work, we were challenged to embark on a large-scale modeling project to thoroughly specify the processes and facts of urban appraisal procedures of a municipality [10]. The urban appraisal procedures handle the requests regarding permits for all kinds of building construction work, land subdivision, and other urban operations within a municipality. The result of this project supported the parameterization of a software system to manage these procedures.

We applied the standard DEMO representations' methodology. However, this project faced an order of magnitude higher than our previous experiences in terms of process complexity. Some numbers illustrate it: 393 tasks across 13 processes, with a total of 693 dependencies (including 280 causal links and 69 waiting links), 506 facts, 441 attributes, and 65 concepts aggregating them. In addition, we identified 567 relationships between the urban appraisal processes and the necessary instructional elements (mostly different kinds of documents/plans).

Therefore, in the early stages of the project, we were already faced with overly cluttered and complex diagrams. The modeler experts had already struggled with this complexity. Furthermore, in model validation meetings with the municipality's officers (including architects and lawyers, who were also accustomed to complexity), they expressed concerns that the diagrams were overly complex, visually cluttered, and hard to interpret. Both groups felt a strong need for cleaner and friendlier representations with simplified and richer semantics [11].

We also found it difficult to maintain different diagrams with some redundancies according to every change identified as needed during the validation stages [10]. Furthermore, the sheer number of symbols, elements, and text within the Fact Model, specified in the General Ontology Specification Language (GOSL) [12], overwhelms users with no prior DEMO experience [11].

To address these limitations and improve the user experience in this modeling project, we started to design and iterate new simplified notations for representations of DEMO's Process and Fact Models, giving special attention to officers' feedback. Simpler and semantically richer representations of the Process Model were proposed in [10]. They included innovations — inspired, in part, by the Business Process Model and Notation (BPMN) standard [13] — and introduced a new representation table, making it much easier to understand the process flow and locate relevant information [10].

New representations for the Fact Model are presented in [11], with a notation that is simpler and easier to understand. These novel representations mix elements of the typical Entity Relationship Model [14] and Relational Schema [15] diagrams, as well as a new table. All these elements specify, in a detailed and structured way, important aspects of the concepts and their attributes.

We argue that these new representations minimize some drawbacks of the original DEMO way of working. However, it does not entirely solve them. For instance, depending on the size and intricacies of the processes, the complexity of the representations may always pose an issue, regardless of how simple the notation is. Still, we need to acknowledge that, if it is oversimplified, the representation might lose its use by failing to convey all the critical information, leading to misleading interpretations, or even no longer be useful in the decision-making process, or the understanding of the process itself. Therefore, a reasonable balance must be achieved.

Awareness of which representations are more effective is crucial for clear communication and understandability of processes and facts [4], [16], [17]. Therefore, a user-centered approach should be considered in the evaluation of these representations to understand the user experience and motivation when using these artifacts.

The informal feedback we received from the officers regarding the final iteration of the new representations was quite positive, regarding the approaches we designed and their results. However, these contributions lacked a formal evaluation.

To fill this gap, we conducted a family of three evaluation studies to explore user perception of both standard DEMO representations (Versions B and D) [1] and newly proposed representations (Versions A and C) [10], [11], conceived to be more user-friendly. In Section III, we provide details of the

Context of the Studies, whereas in Section IV, we introduce a high-level overview of each study along with its main findings. The results of these three comparative evaluations were presented at scientific meetings and reported in [18], [19], [20].

The goal of this paper is to highlight relevant results from each study and draw new and reinforced conclusions across the data collected in all three studies. That is, we pooled participants' data from all samples while globally applying parametric tests and presenting novel analyses. Consequently, we expanded and consolidated the original findings, increasing the accuracy of the statistical analyses to assess the effects and reliability of the results. Our research focuses on stakeholders' perceptions of quality and functional suitability, going into the barely explored area of user experience within DEMO models.

TABLE 1. Hypotheses formulated.

Hypothesis	Description
H_1	Newly proposed representations (Versions A and C) are rated as having higher overall perceived quality and functional suitability.
H_2	Newly proposed representations (Versions A and C) are rated as having higher empirical quality.
H_3	Newly proposed representations (Versions A and C) are rated as having higher social pragmatic quality.
H_4	Newly proposed representations (Versions A and C) are rated as having higher functional suitability.
H_5	Subjects with higher self-reported domain knowledge perceive newly proposed representations (Versions A and C) as having higher empirical quality, social pragmatic quality, and functional suitability.
H_6	Subjects with higher self-reported DEMO knowledge perceive standard DEMO representations (Versions B and D) as having higher empirical quality, social pragmatic quality, and functional suitability.
H_7	Subjects will pick, as preferred information sources during task/facts clarification, the newly proposed representations (Versions A and C) rather than the standard DEMO representations (Versions B and D) or legislation/guidelines.

Before finishing this introduction with the hypotheses we devised, we introduce the structure of this paper, which is organized into nine sections. Following the introduction (I), we present the literature review (II), context of the studies (III), family of studies (IV), method (V), and results (VI). We then discuss findings (VII), present validity threats and limitations (VIII), and conclude by summarizing the contributions and pointing out several future research threads (IX).

A. HYPOTHESES

Seven hypotheses were formulated and tested to compare user perceptions of the newly proposed and standard DEMO representations (see Table 1). These hypotheses investigated the adoption potential of new representations and the influence of self-reported knowledge on user perceptions.

In this family of studies, we hypothesized that participants would perceive the newly proposed representations (Versions A and C) as having higher overall perceived quality and functional suitability (H_1) because of their overall user-friendly design, better intelligibility, and usefulness.

Avila and his colleagues [4] report that there is no consensus in the literature regarding the use of color in process models, as it seems to depend on what type of task needs to be performed. Although, it has also been noted that aesthetics can influence user perception, and the strategic use of color can enhance a model's EQ [16]. Considering the use of color and a more visually appealing design in the newly proposed representations (Versions A and C), we predicted that participants would find them to have a higher empirical quality (EQ) (H_2).

Social pragmatic quality (SPQ) focuses on how well users understand a model's meaning. A clear presentation plays a significant role in it [16]. The reduced clutter and user-friendly appearance in the novel representations (Versions A and C) led us to hypothesize that participants would perceive them as having a higher social pragmatic quality (H_3).

Functional suitability is defined as the capability of a product to provide functions that meet the stated and implied needs of intended users [21]. Considering that the newly proposed versions are claimed to be more accessible, easier to grasp, and closer to users' needs [10], [11], we predicted that Versions A and C would be evaluated as having higher functional suitability (H_4).

Literature suggests that familiarity with a modeling language can positively influence comprehension [16]. Thus, we hypothesized that participants with higher self-reported knowledge of the modeling language (DEMO knowledge), would perceive traditional representations (Versions B and D) as having higher empirical quality, social pragmatic quality, and functional suitability (H_6).

Considering the effect of familiarity with the modeling language, we wondered whether higher self-reported domain knowledge — that is, higher familiarity with the modeled domain — would somehow influence the same variables above. Therefore, we foresaw that those with higher knowledge of modeled concepts should be receptive to less cluttered representations, facilitating an overview of modeling concepts, processes, and facts. As a result, we formulated Hypothesis H_5, predicting that subjects with higher self-reported domain knowledge should pick Versions A and C when prompted to select those with higher empirical quality, social pragmatic quality, and functional suitability. By explicitly hypothesizing the domain knowledge dimension, we also foresaw the possibility of exploring variations in the results across different domains, eventually allowing us to reinforce possible conclusions and mitigate potential external threats to validity regarding generalizability.

Finally, because of the focus on user experience when developing Versions A and C, we hypothesized that participants would find them to be more helpful information sources for tasks/facts clarification compared to traditional DEMO diagrams (Versions B and D) or other information sources such as legislation/guidelines (H_7). Reducing visual clutter in the newly proposed representations is expected to facilitate navigation and understanding of the relationships between tasks and facts. In addition, compared to consulting legislation/guidelines, newer representations, by including systematically structured and dynamic descriptive tables, should be perceived as more effective means for locating desired information.

II. LITERATURE REVIEW

Representations of business models that are comprehensive and successfully communicated are critical for the implementation of effective information systems [6]. Nonetheless, the impact of business models goes beyond its content. Their depictions play a crucial role in how users comprehend them. Research suggests that when compared to content, representation can have an equal or even greater impact on cognitive effectiveness (speed, ease, and accuracy) [22].

According to Zeevat [23], if a logical representation language is used in the formulation of grammar, the relation between syntactic objects and their representations, and between the representation and their meanings, must be expressed as homomorphisms between the relevant algebras. This representation helps to develop grammar by making it easier to understand. In principle, we can eliminate it in favor of homomorphism from syntax to semantics [23].

Several frameworks have been developed to make process modeling effective [4], including GoM (Guidelines of Modeling) [3], 7PMG (Seven Process Modeling Guidelines) [2], the Physics of Notations theory [24], and SEQUAL (Semiotic Quality Framework) [16]. These frameworks set forth user-centric principles such as perceptual discriminability (i.e., the ease with which users can differentiate symbols and understand their meaning) [17]. Despite the existence of guidelines, evaluating the adequacy of business model representations remains a challenge [4], [6], [17]. Different existing evaluations focus largely on the technical assessments conducted by modelers [17]. However, a user-centric approach that includes satisfaction and comprehension is an essential feature of a proper evaluation framework [25]. Research has explored some of these areas [4] but thorough methods and criteria for assessing perceived quality and stakeholder satisfaction within our modeling context are still unavailable [6], [25].

This section will address four areas: 1) a brief approach to the theories of Cognitive Theory of Multimedia Learning and the Cognitive Load Theory (II-A) — to elaborate on the cognitive processes connected with learning; 2) guidelines for process modeling (II-B and II-C) — where we will go over the most popular frameworks in process modeling; 3) evaluation of diagrammatic representations (II-D and II-E) — to analyze the most important models for assessing the perceived quality of representations; and 4) functional suitability (II-F) — looking at how constructs provide functions that meet the stated and implied needs of intended users.

A. COGNITIVE APPROACHES TO LEARNING

The Cognitive Theory of Multimedia Learning (CTML) [26] and the Cognitive Load Theory (CLT) [27] explain the cognitive processes involved in understanding and retaining information.

CTML holds three core concepts [26]:

1) Dual channels - Humans have separate information processing channels for verbal and visual information.
2) Limited capacity - Processing capacity is severely limited.
3) Active processing - Meaningful learning involves selecting relevant material to be processed in working memory, mentally organizing the material into coherent verbal and visual structures, and integrating them with each other and with relevant knowledge activated from long-term memory.

CLT considers the aspects of human cognitive architecture concerned with how people learn, think, and solve problems [27]. This theory endorses that learning will be facilitated by optimizing intrinsic cognitive load and reducing extraneous cognitive load [27].

Therefore, considering humans' limited capacity, materials presented using both words and visuals are easier received, understood, and retained by users [26]. CLT complements this idea, addressing the limitations of working memory and the concept of element interactivity [28], endorsing that, to acquire new information, extraneous cognitive load should be decreased to the minimum [27].

B. GUIDELINES OF MODELING (GOM)

The Guidelines of Modeling (GoM) address the need for design recommendations to improve the quality of information models that go beyond simply following syntactic rules [3]. It aims to improve the quality of information models by considering six aspects: correctness, relevance, economic efficiency, clarity, comparability, and systematic design (see Fig. 1) [3].

The first three concepts (correctness, relevance, and economic efficiency) are recognized as important components of model quality. They ensure that the framework properly reflects reality, presents accurate information, and is constructed with attention to resource efficiency [3]. The other three concepts (clarity, comparability, and systematic design) were considered discretionary but beneficial. In conjunction, they provide the foundation for designing high-quality information models [3].

C. SEVEN PROCESS MODELING GUIDELINES (7PMG)

Mendling [2] proposed 7PMG as a straightforward set of recommendations that specifically focused on the visual

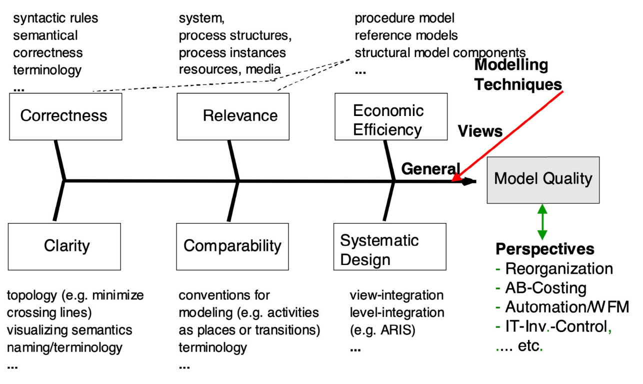


FIGURE 1. The framework of the GoM [3].

representation of process models as opposed to their content. These guidelines aim to improve user perceptions by promoting clarity and reducing complexity. The recommendations are as follows: G1 — Use as few elements in the model as possible. G2 — Minimize the routing paths per element. G3 — Use one start and one end event. G4 — Model as structured as possible. G5 — Avoid OR routing elements. G6 — Use verb-object activity labels. G7 — Decompose model with more than 50 elements.

Although these guidelines can be applied at various points within a process model, it is important to acknowledge the potential interactions and conflicts that may arise, which may undermine representation accuracy [2].

D. PHYSICS OF NOTATIONS THEORY

The Physics of Notations Theory [24] offers a framework specifically designed to evaluate and improve the visual language of modeling languages [24]. This theory proposes nine evidence-based principles that can be employed to assess and improve the effectiveness of the visual notations of modeling languages:

1) Semiotic Clarity - There should be a one-to-one correspondence between semantic constructs and graphical symbols.
2) Perceptual Discriminability - Symbols should be undoubtedly distinguishable.
3) Semantic Transparency - Use symbols whose appearance is evocative.
4) Complexity Management - Includes mechanisms for handling complexity.
5) Cognitive Integration - Includes explicit mechanisms to support integrating information from different diagrams.
6) Visual Expressiveness - Use visual variables' full range and capacities.
7) Dual Coding - Enrich diagrams with textual descriptions.
8) Graphic Economy - Keep the number of different graphical symbols cognitively manageable.
9) Cognitive Fit - Use different visual dialects for different tasks and audiences.

Evaluations of notations regarding these principles often rely on the values of visual variables, that is, the elementary characteristics forming the visual alphabet of diagrammatic notations [17]. However, the Physics of Notations does not offer a comprehensive theory to analyze, evaluate, and improve the cognitive effectiveness of visual modeling languages, even when a set of symbols and semantic constructs are defined [17]. Furthermore, these principles are more focused on a technical evaluation of the representation and not on an assessment by the users of the diagrams regarding their perceived quality and functional suitability. However, these principles can also be used to generate empirically testable predictions.

E. THE SEMIOTIC MODEL QUALITY (SEQUAL)

The SEQUAL framework [29] offers an approach to model quality by focusing on the social and user-centered aspects of modeling activities [16], [29], [30]. It views modeling activities as socially situated (constructivist perspective) where stakeholders engage in dialogue, and their understanding of the domain changes as the model is built [17]. In information systems, evaluating the quality of process models is challenging [29], [30]. Issues like the language's formality, model complexity, and effort required to understand its components can hinder user comprehension [16].

SEQUAL identifies seven quality dimensions that go above technical correctness [16]:

1) Physical - The persistence, currency, and availability of the process model.
2) Empirical - The relationship between the process model and another process model that contains the same statements, which is regarded as better through a different arrangement or layout.
3) Syntactic - The relationship between the process model and the process modeling language.
4) Semantic - The relationship between the process model and the modeling domain. Perceived semantic quality is the parallel relationship between participants' knowledge and interpretation of the process model.
5) Pragmatic - The relationship between the process model and stakeholders' interpretation of the model.
6) Social - The relationship between different process model interpretations.
7) Deontic - The fit between the process models and the modeling goals.

Fig. 2 illustrates SEQUAL framework, namely its main concepts and relationships.

One of SEQUAL's strengths lies in differentiating between the perception and technical aspects of quality, that is, distinguishing the social pragmatic and empirical quality dimensions [29]. Empirical quality (EQ) concerns the technical aspects influencing a model's visual representation [16]. This includes two key sets of variables:

- Planar variables - Encompasses the horizontal and vertical positioning of elements within the model;

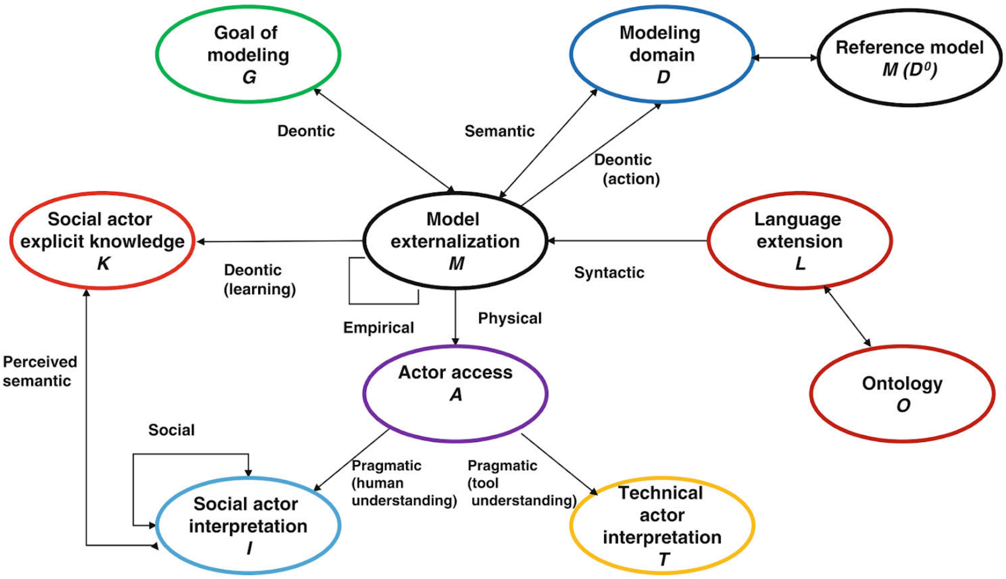


FIGURE 2. The SEQUAL framework [16].

- Retinal variables - Related to the visual properties of the elements, including shape, size, color, brightness, orientation, and texture.

In addition to these core variables, other dimensions contribute to a model's EQ [16]:

- Visual emphasis - Includes techniques such as solidity, pattern variations, foreground/background contrast, change (movement), position, and connectivity;
- Color coding - While some argue for avoiding color-coded meaning, its effectiveness as a differentiator is recognized. That is, a well-defined color scheme can improve the EQ of a model; and
- Aesthetics - Plays a (subjective) role in user perception and EQ, for example. Existing guidelines for graph aesthetics can be used to improve a model's technical representation.

Besides, familiarity with a diagram's style can positively influence its perceived EQ [16].

The SEQUAL framework defines social pragmatic quality (SPQ) as the degree to which a model is understood by its intended audience [16], [29]. This concept is multifaceted, distinguishing between perceived social pragmatic quality (how well human stakeholders understand the model's meaning) and technical pragmatic quality (a model's compatibility with software tools that might interpret it) [29].

While evaluating these variables, one can assess whether the diagram has been understood and who has understood (the relevant parts of) it [16]. This ensures an effective communication. This framework further breaks down the SPQ into four key dimensions influencing user understanding: language perception, content relevance, structured analysis, and behavior experience. By understanding these factors, model creators can develop representations suited to their audiences' specific needs and knowledge bases, promoting better comprehension and communication. While some individuals are familiar with formal languages and, consequently, a formal model guarantees understanding, others find a mix of formal and informal statements more discernible, strengthening their language perception [16]. Behavior experience considers the user's previous experience with similar models and how it influences their interpretation of a current one. As could be expected, participants with previous experience in modeling find it easier to understand new representations [16].

Krogstie's framework includes a detailed set of quality dimensions, but no evaluation metrics, serving as a reminder that model quality is more than just technical accuracy; it should lead to effective communication with the intended audience.

F. FUNCTIONAL SUITABILITY

In software quality models, functional suitability refers to a product's ability to meet the stated and implied needs of intended users under specific conditions [21]. Furthermore, functional suitability implies the fulfillment of a product's functional specifications [21]. In this area of software quality, a product is viewed as a broad concept that encompasses software, data, hardware, and other ICT-related artifacts [21].

Within the evaluation of label types in process modeling, literature presents the concept of perceived usefulness. It is defined as the degree to which an individual believes that a label enhances their understanding of the process being modeled [31]. A study by Mendling and colleagues [31] investigated the interaction effect between perceived usefulness, domain knowledge, and familiarity with the notation.

Yet, their findings did not reveal a statistically significant interaction effect [31].

Because we are working with model representations frequently used as an important source for Information Systems/software development, we adopted the concept of functional suitability in our applied research work.

Functional suitability is divided into (a) functional completeness (capability of a product to provide a set of functions that covers all the specified tasks and intended users' objectives), (b) functional correctness (capability of a product to provide accurate results when used by intended users), and (c) functional appropriateness (capability of a product to provide functions that facilitate the accomplishment of specified tasks and objectives) [21]. A comprehensive evaluation of an artifact is key to ensuring its value for stakeholders [21]. This includes assessing the artifacts' characteristics, as well as their impact on stakeholders [21]. In general, users tend to select artifacts that provide the information required to perform their tasks, which they classify as more suitable [21].

The ISO/IEC 25020:2019 standard, part of the SQuaRE (Software Quality Requirements and Evaluation) series, defines the Quality Measurement Framework [32]. It provides a reference model for software product quality measurement, including definitions of quality measures and practical guidance for their application. This standard describes three types of quality measures: internal, external, and quality-in-use.

Complementing this, ISO/IEC 25040:2024 defines the Evaluation Process [33], offering requirements, recommendations, and guidelines for evaluating software products—whether by independent evaluators, acquirers, or developers. The questions used in this study to assess the functional suitability of the DEMO Process and Fact Model representations (see Section V) are aligned with the principles and terminology of these standards.

III. CONTEXT OF THE STUDIES

We developed a family of studies to understand which diagrammatical representations users would prefer. We used traditional DEMO diagrams [1] and newly proposed representations for both the Process Model (Version B and A, respectively) and the Fact Model (Version D and C, respectively) [10], [11]. The artifacts of the newly proposed representations were developed following Design Science Research Cycles [34]. The artifact design was refined through several iterations, based on feedback from a highly experienced team of domain experts. The resulting artifacts were considered complete, simple, understandable, easy to use, and having the potential to aid officers in their daily tasks [10], [11]. It is worth clarifying that we follow the Model Universe (MU) theory approach, which advocates that models are at a subjective level while representations are objective [1]. Representations are concrete artifacts which are an expression of the mental image of models. They can be represented by a set of symbols, diagrams, tables, or any other type of structured written information [1]. In this section, we describe the main concepts behind the diagrammatical representations used in this family of studies and then briefly present Versions A, B, C, and D.

A. PSI THEORY AND DEMO MODELS

PSI-theory [1] is the main basis of the Design and Engineering Methodology for Organizations (DEMO). It frames organizations into three layers: O-organization (O from original), I-organization (I from informational), and D-organization (D from documental) [1]. These interconnected layers form the *essential model* of an enterprise, which abstracts unnecessary details, captures the essence of an organization's operations, and facilitates the management of its complexity [1]. It is a high-level abstraction, independent of specific implementations, and focuses on the key elements crucial for intellectual manageability [1]. The essential model resulting from DEMO is significantly smaller, often less than 5%, compared to other traditional organizational modeling approaches [1].

According to PSI-theory, an enterprise consists of a network of actors that enter into and comply with commitments with each other, known as coordination acts (e.g., a building permit is requested), regarding a production fact (e.g., a building permit is issued) [1]. These commitments are brought about by actors in acts, the basic units of action, and they follow a pattern named the Complete Transaction Pattern (CTP), consisting of nineteen transaction steps, eighteen of them coordination acts revolving around one production act/fact [1]. This includes the normal happy flow of a transaction: request, promise, execute, declare, and accept. The discussion states: decline, reject, and the different kinds of cancellations (also called revocations) are also included [1]. The working principle of PSI-theory for an enterprise is that actors constantly check if there are acts they have to deal with or respond to in their actor cycle [1].

The DEMO standard prescribes a set of models and diagrams to represent an organization. These are the Cooperation Model (CM), the Action Model (AM), the Process Model (PM), and the Fact Model (FM), which are interconnected and specify diverse elements in a neutral and technology-independent manner [1]. The Cooperation Model (CM) specifies the construction of an organization and its transaction types. It specifies the transactor roles (the elements) and the coordination structures (the influencing relationships) between them [1]. The Action Model (AM) is the ontological model of an organization's operation, that is, the manifestation of its construction over time. Every internal actor role is associated with a set of action rules that guide them in performing coordination acts. These action rules are represented in the Action Rule Specifications [1]. Since the CM and AM are out of the scope of our work, this very brief introduction to them suffices. The Process Model (PM) bridges Cooperation and Action models by outlining coordination, which triggers different transactions according to the actors' actions. It contains, for all transactions, the transaction steps and the applicable occurrence laws (response and

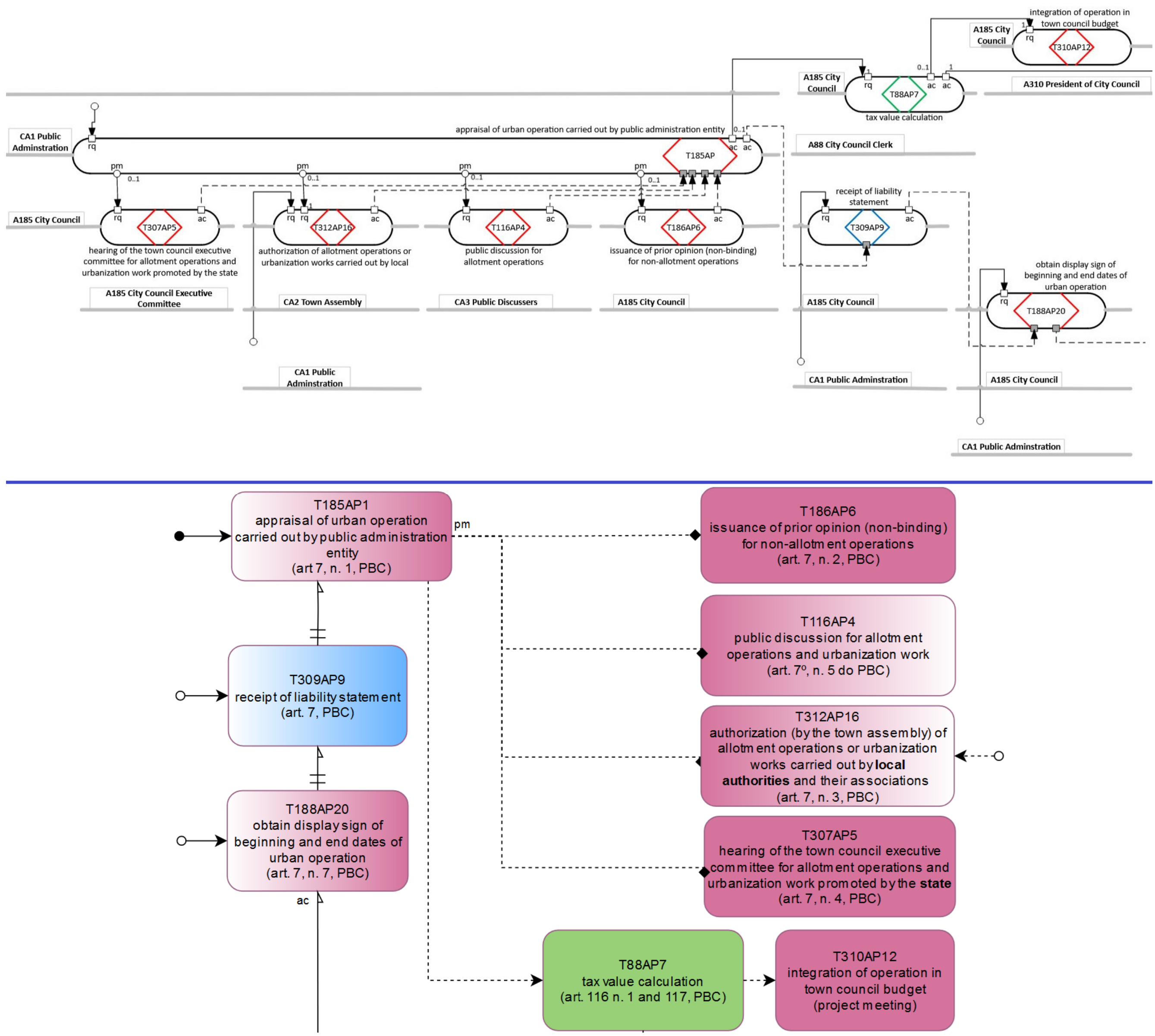


FIGURE 3. Process structure diagram (Version B) on the top, and process diagram (Version A) on the bottom [10].

wait links), including cardinalities of occurrences [1]. For example, a public administration entity starts the appraisal of an urban operation (request transaction step), which the municipality commits (promise step) to assess. When the appraisal is promised, the municipality, in turn, might trigger (request step) other procedures: an issuance of prior opinion; a public discussion; an authorization by the municipal assembly of the land subdivision operations; and a hearing of the municipal council executives. All of these procedures are optional, but according to certain rules and the type of urban operation, at least one or some of them must be completed, so that the result of the appraisal of the urban operation is approved or denied (execution step). Then, this result must be communicated (declare step) to the requesting entity. This entity might then choose to accept (accept step) or refuse (reject step) such decision. This process is represented in both the standard and newly proposed versions of the diagrams in Fig. 3). The Fact Model (FM) focuses on an organization's products/services and their information/attributes. It contains entity, value, property, and attribute types, and the existence laws that apply [1]. It also expresses event types and their occurrence rules [1]. In our example, an application for an urban operation has multiple Application Deliverables. Each of these deliverables has, as attributes: a state which has to be selected from a predefined list of possible values (e.g.: initial, change, enhancement, etc.); a reference to a specific type of application deliverable (e.g., drawing, urban project, liability statement); and a justification that can be a document or simple textual input. Both representations can be observed in Fig. 6.

The family of studies reported in this paper addresses all these representations, which are explained in more detail in the following sections.

B. MU THEORY AND GENERAL ONTOLOGY SPECIFICATION LANGUAGE

One of the foundations of DEMO is the Model Universe (MU) theory [12], which addresses the concepts of models and representations. Various models and diagrams used in DEMO rely on the concepts and principles of the MU theory to ensure that they are formally correct and meaningful representations of the enterprise they depict [1]. To formally convey the rationale of a conceptual schema into symbols and constructs, the MU theory relies on the General Ontology Specification Language (GOSL) proposed in [12]. It is used to formally represent not only the Object and Fact Diagram of DEMO's FM, but also all metamodels of all DEMO models, including of the FM itself. However, as pointed out in our previous research [11], the GOSL notation for the FM, in its current status, is a far from ideal choice for drawing representations that are easily understood and productively discussed by a full range of stakeholders, regardless of their technical skills, experience, or background.

C. STANDARD REPRESENTATION OF THE PROCESS MODEL

The PM of an organization is the ontological model of the state space and the transition space of its coordination world [1].

A PM is represented in a Process Structure Diagram (PSD), optionally supplemented by Transaction Process Diagrams and a Create Use Table [1]. The transaction kind shapes are stretched horizontally, thus becoming sausage-like shapes [1]. At the top of Fig. 3, a partial example of the Process Structure Diagram in standard notation (Version B) is shown. This illustrates the same part of the process, also depicted in the newly proposed representation (Version A), at the bottom of Fig. 3. This example falls within the scope of urban operations in municipalities and depicts the procedure for requests of construction work carried out by another public entity (e.g., a regional government).

A PSD shows the inter-transaction occurrence laws, that is, the laws held between transactions of different kinds [1]. They are expressed in two kinds of process links: response and wait links [1]. As an example of a response link, we have the case of the request of the transaction *tax value calculation*, as a response to the transaction *appraisal of urban operation carried out by public administration entity* having been accepted. An example of a waiting link is that the execution step of the transaction *appraisal of urban operation carried out by public administration entity* (the final decision) has to wait for the final step (accepted) of the transaction *hearing of the town council executive committee*.

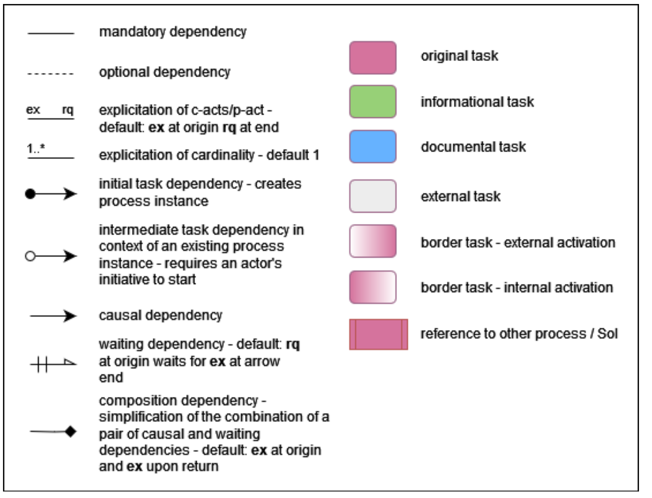


FIGURE 4. Process diagram legend [10].

The bold gray lines behind the sausage shapes separate the responsibility area of the initiator (above the line) from the executor (below the line) [1]. Following our example, the *Public Administration* role is responsible for requests of transaction *appraisal of urban operation carried out by public administration entity*, and the *City Council* role is the one responsible for the execution (decision).

D. NEWLY PROPOSED REPRESENTATIONS OF THE PROCESS MODEL

Recent work by our team [10] proposed new representations to depict DEMO's Process Model for the reasons discussed in Section I. These representations extend the standard by incorporating elements from Cooperation and Action Models, as detailed in [10].

The new approach introduces the following two representations [10]:

- Process Diagram - More concise than the standard PSD, making it easier for both modelers and stakeholders to understand. Each task (simpler name, used as a synonym of transaction) symbol includes a reference to its information source to facilitate the tracking of detailed information. Fig. 4 illustrates this notation.
- Transaction Description Table (TDT) - This table complements the Process Diagram by providing detailed structured information in a textual and hyperlinked manner. It includes actor roles, organizational functions, and specific rules (Fig. 5) [10]. The TDT achieves better fulfillment of the same function as the standard DEMO Action Model owing to the perceived limitations in clarity and connection to real-world operations of the latter [10].

A key improvement is the development of a simpler notation for the Process Diagram. This new approach offers a clear separation of concerns regarding process composition, task causation, and task waiting. Regarding connectors, diamonds represent the compositional perspective, arrows

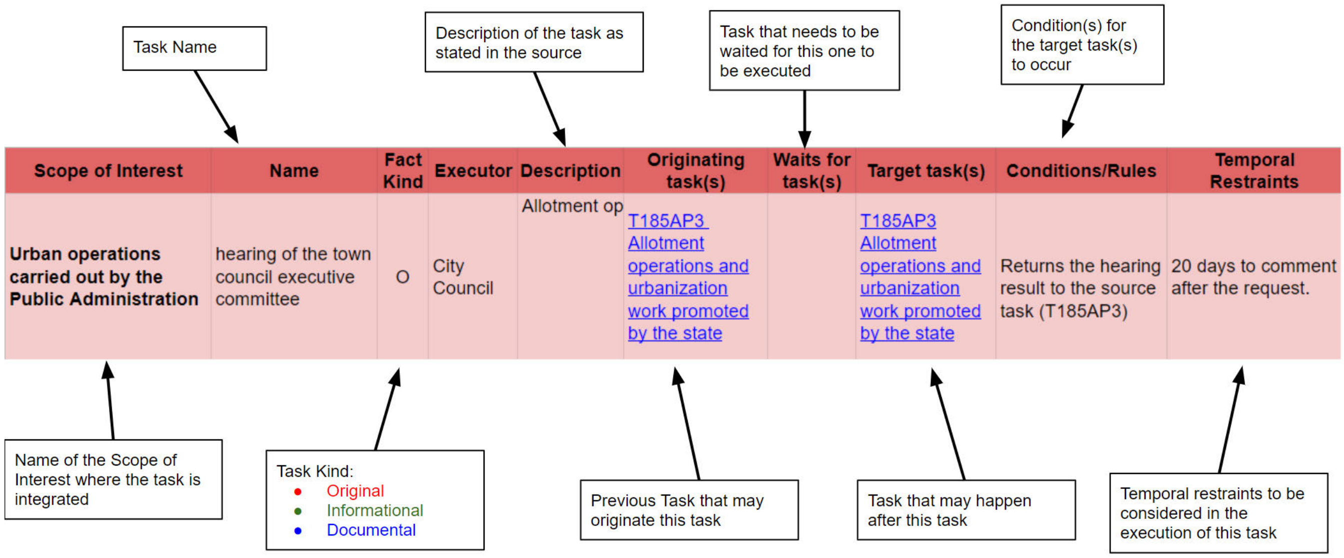


FIGURE 5. Transaction description table [10].

depict causal relationships between tasks, and double-crossed lines indicate waiting dependencies. The color gradients on border transactions facilitate rapid comprehension of task initiation and execution. These gradients distinguish between tasks initiated by internal actors and those executed by external actors (and vice versa). The solid lines represent the mandatory tasks, whereas the dashed lines indicate the optional tasks. The cardinality (frequency), is represented using number indicators at the endings of connectors, such as *1..**. For the most common case, where the cardinality is *1..1*, its representation is omitted as to not clutter the representation.

The advantages of the new representations are pointed out [10]:

- Reduced diagram complexity - The new approach simplifies diagrams through improved symbol design, fewer connectors, and the omission of explicit actor roles, which avoids overloading the visual representation with unnecessary details, while more information about the tasks is conveyed in the TDT.
- Enhanced readability - The use of color, gradients, and specific symbols for related processes improves comprehension of process boundaries and the nature of the tasks
- In-depth insights - The new notation offers a deeper understanding of modeled processes by incorporating elements from other DEMO models.
- Complete and structured task information - The TDT addresses the information overload in the standard diagram by providing a dedicated space for detailed textual information on actors, roles, task descriptions, and different rules that control task flow and interaction with other tasks.

Regarding the Process Model, our family of studies compares the newly proposed representation, labeled Version A, with the standard DEMO notation, labeled Version B. Both are illustrated, respectively, at the bottom and top of Fig. 3.

E. STANDARD REPRESENTATION OF THE FACT MODEL

The Fact Model (FM) is an ontological model of the state space and transition space of an organization's production world [1]. Regarding the state space, the FM contains entity types, value types, property types, and attribute types that are relevant to the modeled organization and existing laws that apply. Concerning the state space, an FM is the conceptual schema of the production world of the modeled organization. Regarding the transition space, an FM contains the event types and the occurrence laws that are applied. The FM of an organization connects its Cooperation Model and Action Model, as far as production is concerned [1]. The standard representation can be observed in the top part of Fig. 6.

An FM is expressed in an Object Fact Diagram (OFD) supplemented by (textual) Derived Fact Specifications (DFS). An example of a DFS is: *urban construction tax = location coefficient * kind of use coefficient * infrastructure costs * value per square meter * total construction area * adjustment factor*.

The current standard representation of the OFD in GOSL (Fig. 6, top) is considered to have too many elements and text, making it difficult for officers without experience in DEMO to interpret it [11]. The GOSL syntax is as follows:

- Roundangles (without diamonds inside) - Represent entity classes, thus, they are extensions of entity types. In these rectangles, we can also represent the accepted domain and the range of accepted values (e.g., "assessment of the urban project {DOCUMENT}" of an entity "URBAN PROJECT"). When the name in the round rectangle is between { }, it represents a Value Type (e.g., "{TYPE OF URBAN PROJECT}").

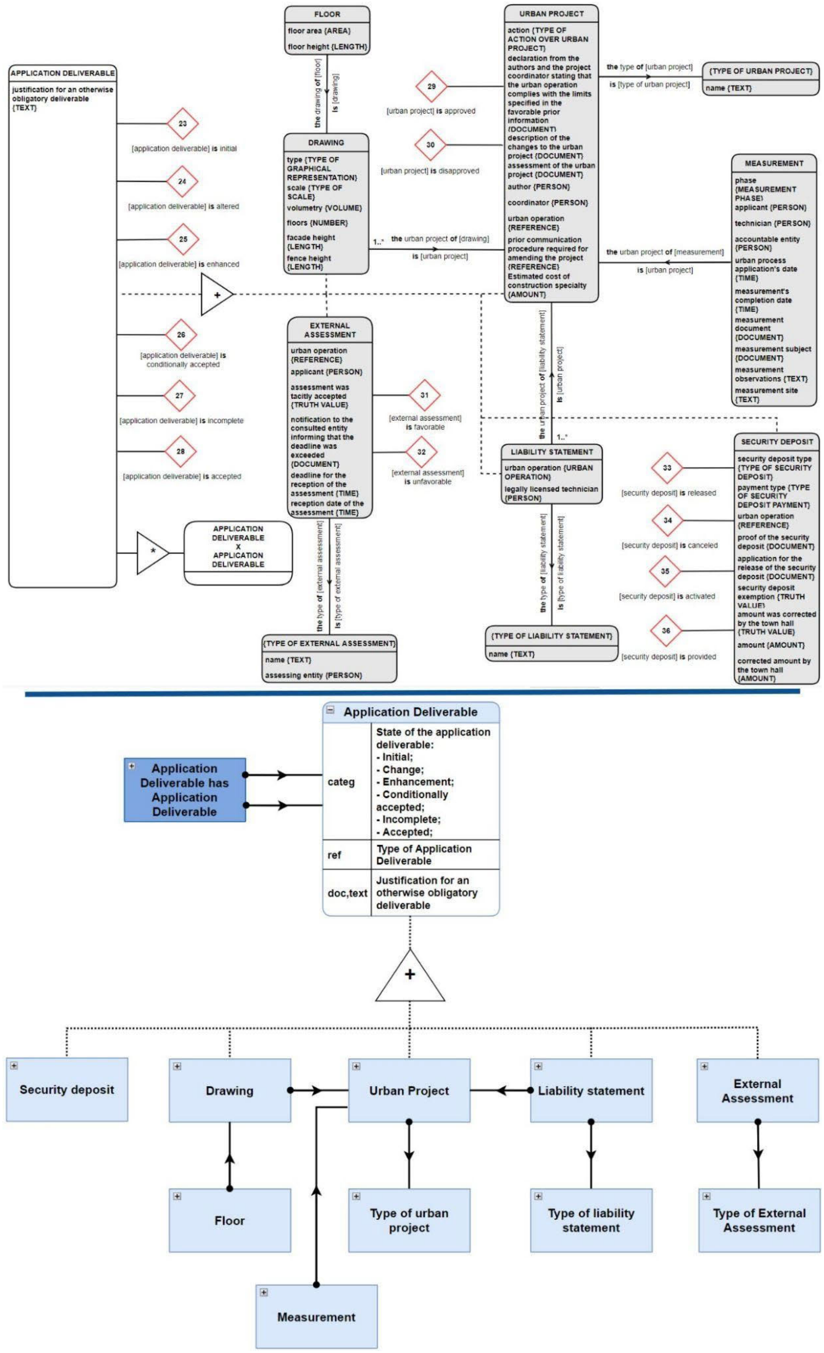


FIGURE 6. **Object fact diagram (top - Version D) and concept and relationships diagram (bottom - Version C) [11].**

- Lines with an arrow in the middle - Can be considered collections of separate mappings from elements.
- Cardinality - The ends of connectors generally mean 1. If a range string is present (for example 1..*) it denotes the minimum and maximum cardinalities applied to that end.
- Diamonds - Represent event types (i.e., the p-facts of transactions) that apply to each class such as [urban project] is approved.
- Triangles with a '+' sign inside - Correspond to generalizations like "APLICATION DELIVERABLES" is a generalization for "DRAWING", "URBAN PROJECT", "EXTERNAL ASSESSMENT", "LIABILITY STATEMENT" and "SECURITY DEPOSIT".
- Triangles with a '*' sign inside - Correspond to aggregations for example the aggregation of "APLICATION DELIVERABLES".

Concept	Attribute name	Value Type	Referenced Concept / Category values	Description	Source - task 1	Step	Task 1 (creates or updates the attribute)
Concepts of Application Deliverable Management							
Application Deliverable	**Application Deliverable**			The "Deliverable" concept specifies the instances that are handed over to the Municipality's services as part of the management of deliverables procedure. Supports the preliminary analysis task by making it possible to discriminate which deliverables were effectivelly delivered.	RJUE - Art. 11	Ex	T197CP5 Preliminary analysis of urban planning / building process (preliminary assessment)
Application Deliverable	State of the application deliverable	category	Change	This attribute holds the the state of deliverables (e.g. if it is the initial handing over of the element, if it is an enhacement or if this element has already been accepted): Possible options for the state of the deliverable are presented as follows: - Initial: the initial delivery; - Change: a change (potentially unsolicited) of a given deliverable; - Improvement: a new iteration of a deliverable following an enhacement request; - Conditionally Accepted: the deliverable was accepted minus some impending changes which are yet to be delivered; - Accept: an accepted deliverable; - Incomplete: the deliverable has flaws in its components;	RJUE - Art. 11	Ex	T197CP5 Preliminary analysis of urban planning / building process (preliminary assessment)
Application Deliverable	Justification for an otherwise obligatory deliverable	document		According to the Decree No. 113/2015 of April 22nd there may be deliverables which are exceptionally considered unnecessary to hand over for a particular urban operation.	RJUE - Art. 11	Ex	T197CP5 Preliminary analysis of urban planning / building process (preliminary assessment)
Application Deliverable	Type of deliverable	reference	Type of Application Deliverable	From the analysis of Decree No. 113/2015 of April 22nd, a survey of types of deliverables was promoted. Given that upon delivery, there is a need to maintain which types of deliverables were actually handed over, the correspondence between what was delivered and the type of deliverable is implemented in this attribute.	RJUE - Art. 11	Ex	T197CP5 Preliminary analysis of urban planning / building process (preliminary assessment)

FIGURE 7. Fact description table [11].

- Dashed lines with an arrow at the end - Correspond to specializations.
- Roundangles with diamonds inside - Are also specializations of entities but in the form of event types.

Similar to process model representations, our family of studies compares the newly proposed representation (Version C) with the standard DEMO notation (Version D). To exemplify these notations, Fig. 6 shows an example from the scope of urban operations. It shows representations of the same modeled concepts, allowing for a direct visual comparison (top: standard notation, bottom: newly proposed representation).

F. NEWLY PROPOSED REPRESENTATIONS OF THE FACT MODEL

Standard DEMO Fact Model representations often pose comprehension challenges for stakeholders unfamiliar with the notation [11]. To address this, we devised the following three representations, which follow a layered approach with increasing detail [11]:

- Concept and Relationships Diagram (CRD) - This diagram offers a high-level view of the domain's core concepts and their relationships, promoting stakeholder understanding without overwhelming them with detailed information on the attributes (Fig. 6, bottom). Nevertheless, it is possible to quickly inspect the attributes of a concept because symbols are expandable/collapsible. An example of this inspection is shown in Fig. 6, regarding the Application Deliverable concept.
- Concept Attribute Diagram (CAD) - This diagram can be considered as an extension of the CRD; it is basically the same diagram but with all concepts expanded, presenting their associated attributes in a one-attribute-per-line format. After CRD analysis/validation, this diagram allows for more detailed discussions of the concept details.
- Fact Description Table (FDT) - It (Fig. 7) offers the most detailed level of information in a manner similar to the TDT of the PM. It includes all concepts and attributes presented in the previous diagrams, while providing detailed descriptions for each. It allows full traceability by referencing the source (e.g., legal document) that justifies the existence of the attribute and identifies the transactions responsible for creating or updating its values and in which step of the transaction.

In this family of studies, we compared the newly proposed representation (Version C) with the standard DEMO notation (Version D). In Fig. 6, one can see both representations of the same modeled concepts (top: standard notation, bottom: newly proposed representation).

G. LIMITATIONS OF THE NEW REPRESENTATIONS

Even though the newly proposed representations improve the clarity and reduce the cognitive complexity associated with the traditional DEMO models, there are some limitations that should be pointed out.

In the new process model representations, we chose, by design, to explicitly exclude actor role information (initiators and executors), which appear in the standard representations. This design choice was made based on feedback from the stakeholders of the municipality project, as they preferred a cleaner visual, focused on the process flow. However, this may reduce the transparency in role responsibility, an important key point in the DEMO methodology. To address this, a modelling tool we are developing [35] reveals actor roles via hoverable tooltips, allowing users to access this information as needed without overloading their view. The tradeoff of this option is that it cannot be printed, so we are also considering an additional general toggle configuration to show or hide actor role information in the symbols, so it can be shown in printed diagrams.

The use of colors is a traditional feature of DEMO representations, which uses red to represent original tasks/facts dimension, green for derived or computed tasks/facts dimension, and blue for the documental tasks/facts dimension. Although helpful for the majority, this can introduce a serious accessibility barrier for individuals with color blindness. As a solution for this issue, we plan to add to our modelling tool, dynamic adaptive color schemes that can be changed on the fly based on the different color blindness patterns that a user may have specified in his or her profile.

Regarding the Fact Model representations, a limitation exists in our current symbols and use of modelling tool. When the entity symbols are expanded and collapsed, the layout can lose its proper alignment, and require manual repositioning of the symbol positions in the vicinity and manually rotating the cardinality arrows. This can be especially cumbersome in complex diagrams. An improved version of the modeling tool will support automatic layout adjustment. In a future version of the symbols and tool, we also plan to replace the UML-style centered arrows with the ER-style one-sided arrows to better reflect the meaning of cardinalities, simplify the diagram interpretation, and facilitate the diagramming process.

Another limitation in all representations of both models is navigability and accessibility. As representations increase in their size and complexity, users often need to cross-reference between the different process and fact models and the associated transaction or fact description tables. It is hardly plausible that then can view both at the same time in a single screen, especially if a small one. To improve this, hyperlinking mechanisms were added, to allow easier navigation between related elements. We can also collapse related rows in groups, have automatic identifier columns, and use contextual filters to simplify and improve table navigation. These help to minimize the disconnection between the different representations, but it is only natural that in complex models, some user disorientation might still persist.

Another issue is replication of data of identifiers and names of model elements in the representations. This occurs between the diagrams and respective tables — as they currently are based in two different applications (DISME for diagrams and Google Spreadsheet and Apps Script for the tables) — and also in the diagrams themselves (sometimes we need to represent the same model element in different places of the diagram to facilitate layout and interpretation). Currently, one needs to manually replicate this data from one to the other or vice versa. This implies a risk of errors and mismatching when changes are made. This issue is also being addressed in current developments of our modelling tool, where the table functionalities currently implemented in Google Spreadsheets are being implemented in our tool. Since all data will come from the same source, any change will be immediately visible in all representations, either tables or diagram symbols.

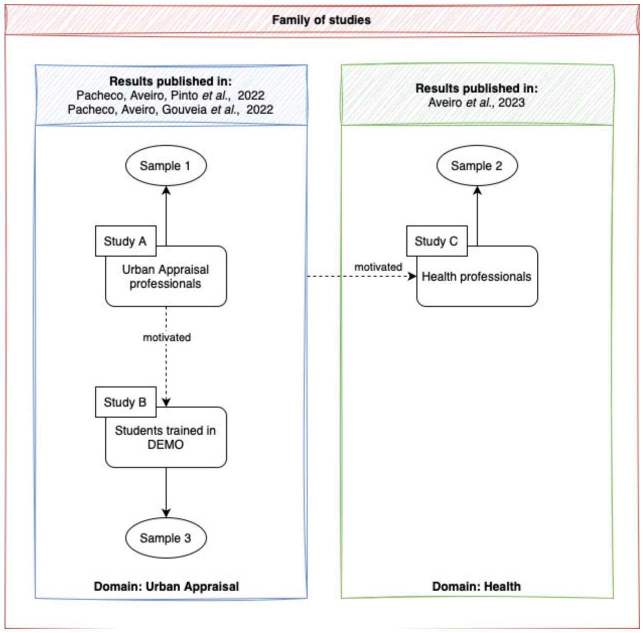


FIGURE 8. Family of studies.

IV. FAMILY OF STUDIES

A family of studies is a set of related studies that focus on the same research question or hypotheses [36]. Fig. 8 depicts our family of studies. Each study tested the same basic hypotheses while addressing different business domains and exploring participants' diverse modeling knowledge.

The results of Study A, indicating that urban appraisal officers clearly preferred the new representations, lead us to explore, in Study B, whether participants experienced in the DEMO language standard would also show such a strong preference for the newly proposed version, rather than the notation that they already knew. After the results were obtained in the domain of Urban Appraisal, we wondered whether Study A might have been biased by the background of the participants (half of them architects), who could have a particular vision over graphical representations. We aimed to explore whether participants from different professional settings would yield similar results, reinforcing our hypothesis in a different domain. Thus, Study C was performed on the health domain.

To facilitate the new and combined analysis of the results presented in this paper, we grouped the participants who did not have DEMO training into Samples 1 (Study A) and 2 (Study C). Sample 3 comprised students trained in DEMO (Study B). Studies A and B have already presented combined analyses, so they can also be considered a (simple) family of studies [18], [19].

In addition, in this paper, we clarified the meaning of the variable *functionality* used in Studies A, B, and C, regarded by the reviewers of [20] (reporting on Study C), as vague. We ground this variable on the family of the norms ISO 25000 — Systems and Software Engineering — Systems and Software Quality Requirements and Evaluation (SQuaRE). Namely, the concept of functional suitability presented above in II is completely aligned with the subject and purpose of the respective questions in the questionnaire.

A. STUDY A-URBAN APPRAISAL PROFESSIONALS

This family of studies emerged from a large-scale modeling project analyzing the Portuguese Building Code regarding municipal urban appraisal requirements and procedures in the municipality of Funchal, Portugal. The urban appraisal procedures of a municipality handle the requests regarding permits for all kinds of building construction work, land subdivision, and other urban operations within its jurisdiction. The municipality was in the process of acquiring a new software system to manage these processes, and needed a detailed study and models of their current processes to support the successful implementation and parametrization of the new system.

Following our experience with previous projects, we applied DEMO standard representations, as it allows us to focus on the essence of the organization, offering a concise and comprehensive representation of the organization [37], while decreasing complexity compared to more traditional approaches. In [1], this reduction was quantified as approximately 95%, a claim supported by our findings in [38], where we obtained a reduction of 90% in model complexity.

Compared to other more widespread approaches, the traditional DEMO diagrams allow a more manageable and clearer representation of processes and facts. However, when modeling the Portuguese Building Code, we found that representations were still too cluttered and complex, making it difficult for both modeling experts and municipal officers to interpret them. This modeling project involved the representation of 393 tasks across 13 processes with hundreds of dependencies between them, as well as the spread of more than 500 facts between concepts and attributes.

Therefore, we identified a need for clearer and more user-friendly representations. To address this issue, we developed and iterated simplified notations for DEMO's Process and Fact Models, incorporating feedback from officers and our modeling experience. These innovations, inspired by standards such as BPMN, evolved into new, more understandable representations. They were reported separately for the Process Model in [10] and the Fact Model in [11]. The informal feedback gathered by officers and the administration was quite positive. Nevertheless, the evolution proposed for DEMO representations lacked formal validation.

Thus, after the end of the modeling project, we recruited a group of professionals from the municipality with knowledge of the modeled processes (Sample 1: $N = 8$) to comparatively evaluate the traditional DEMO representations and the new version we had developed. This gave origin to Study A, where we found, regarding the Process Model, that the perceived quality and functional suitability of the newly proposed representations (Version A) were significantly better evaluated than the standard ones (Version B) [19]. The same was true for representations under the Fact Model [18], showing that Version C is easily understood and can be productively discussed by a full range of stakeholders, independently of their technical skills, experience, or background.

These results led us to question whether subjects trained in the traditional DEMO standard would evaluate these representations differently, which motivated us to develop Study B.

B. STUDY B-STUDENTS

To complement the findings of Study A, we recruited, for Study B, a group of students (graduate and master levels) trained in the standard modeling language in DEMO (Sample 3: $N = 14$).

This study supports the findings of previous research. Participants experienced in DEMO evaluated the quality and functional suitability of the newly proposed representations more positively than the standard representations [18], [19]. Previous researchers have mentioned that familiarity with a diagram positively influences the evaluation of EQ [16]. However, this was not the case in this sample. Remarkably, in this study, the versions that scored a higher level of quality and functional suitability were Versions A and C, which they were not familiar with.

When reporting the results of Study B, we presented both individual analyses and aggregated analyses combining data from Studies A and B. These combined analyses suggested that Versions A (new PM) and C (new FM) were perceived as having higher EQ, SPQ, and functional suitability. The same pattern emerged when each study was analyzed separately.

Interestingly, a significant difference was observed in the evaluation of Versions B (standard PM) and D (standard FM) when comparing responses from students trained in traditional DEMO representations to those from untrained professionals. Specifically, the trained group rated the traditional representations significantly higher. This aligns with Krogstie's claim that familiarity positively influences the evaluation of EQ [16].

However, as already mentioned above and will be discussed in detail in the Results section, the overall hypothesis regarding familiarity was surprisingly contradicted: trained professionals ultimately rated the new representations (Versions A and C) higher than the standard ones (Versions B and D).

Since there could be some bias stemming from the fact that Studies A and B examined the same domain (urban appraisal), and to increase the reliability of the results, we replicated Study A in a different domain (healthcare) in the form of Study C.

C. STUDY C-HEALTHCARE PROFESSIONALS

To understand how professionals with different backgrounds would perceive the quality and functional suitability of the representations of the Process and Fact Models, we recruited a group of healthcare professionals in the field of psychology (Sample 2: $N = 10$).

In Study C, there was a clear preference for the newly proposed representations, both in terms of the Process and the Fact Model [20]. The publication of Study C was accepted as a short paper; thus, for space reasons, we could not include all our analyses and hypotheses in that publication.

To consolidate the results from the individual studies, reinforce conclusions, and increase their reliability, we pooled all three samples and conducted the combined analysis reported in this paper, consolidating the previous three studies in a structured and systematic family of studies.

V. METHOD

One of the best ways to identify real problems affecting user performance and preference is through user testing [39]. Therefore, to evaluate the perceived quality and functional suitability of DEMO's main representations, we present a family of three evaluation studies using mixed research methods ($N = 32$) [40]. The individual results of these evaluations are presented in [18], [19], and [20], while the current paper leverages the pooled data with novel analyses that expand upon the original findings using parametric tests, which are more robust than those used in previous studies (Wilcoxon rank-sum test). The dataset created for the combined analyses is available from IEEE Dataport [41].

Our family of studies employed a mixed-methods approach that utilized both quantitative and qualitative data collection methods. In this section, we describe the participants, materials, and procedures used to collect data for the three studies.

A. PARTICIPANTS

To evaluate the perceived quality and functional suitability of the newly proposed representations comparatively, we recruited participants considering the following:

- Domain knowledge - We included participants who were familiar with the modeled domain. This ensured that they could more easily understand the real-world context represented by the models and provide an informed evaluation capability regarding the quality and functionality of the representations. Sample 1 comprised officers from a local municipality who volunteered to participate in the study. For Sample 2, we recruited healthcare professionals by email invitations sent to healthcare researchers/professionals employed by our research institute and university, as well as to contacts in our professional network.
- DEMO knowledge - We recruited students who were trained in standard DEMO representations. For Sample 3, the researchers addressed a class of enterprise engineering students, offering them the possibility to stay after the lesson was finished, to participate in the study.

A brief characterization of the three samples follows:

- Sample 1 ($N = 8$) - Professionals in the area of urban appraisal with domain knowledge of the modeled concepts (six females and two males). Age range: 33-54 years ($M_{age} = 45$).
- Sample 2 ($N = 10$) - Healthcare professionals with domain knowledge of the modeled concepts (nine females and one male). Age range: 28-58 years ($M_{age} = 34$).
- Sample 3 ($N = 14$) - Students trained in the DEMO modeling language and representations (five females and nine males). Age range: 20-43 years ($M_{age} = 24$).

The total sample ($N = 32$) consisted of more females (62.5%) than males (37.5%) and had a mean age of 32 years (age range: 20-58 years). The participants had heterogeneous backgrounds (see Table 2 and Fig. 9).

TABLE 2. Characterization of the sample.

	Background			
	Architecture	Human/Social Sciences	Law	Other
Sample 1	4	0	3	1
Sample 2	0	10	0	0
Sample 3	0	3	0	11
N	4	13	3	12

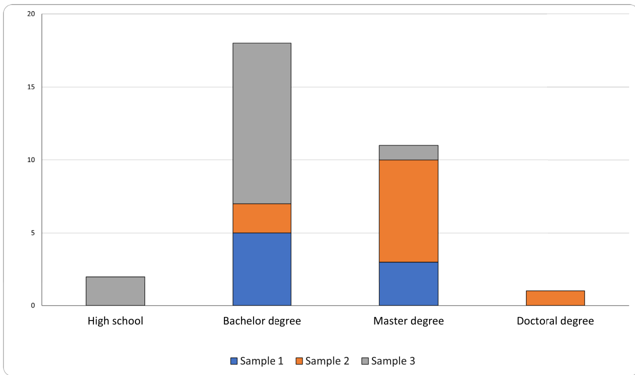


FIGURE 9. Scholar levels of the participants.

B. MATERIALS AND PROCEDURE

The questionnaire used in this family of studies is available online.[1]

1) QUESTIONNAIRE DESIGN

Functional suitability is defined as the capability of a product to provide functions that meet the stated and implied needs of intended users, including the conformity of the products' functional specifications [21]. To evaluate the functional suitability of the representations (designated in a broader way just as Functionality, in the initial studies), the questionnaire included four questions [19]. One question (''is it functional?'') using a six-point scale ranging from 1 = *strongly disagree* to 6 = *strongly agree*, and three questions where subjects were forced to choose which version (A or B; C or D) was considered more suitable (e.g., which version is ''more suitable to support the execution of your tasks'').

In the evaluation of functional suitability, participants completed a scenario-based assessment for each model. Concerning Process Model representation, participants imagined encountering a situation in which they were unsure of the appropriate recipient (individual or department) to forward a particular process during their daily duties. Using a 6-point scale (1 = *definitely not*, 6 = *definitely yes*), they rated their preference for three information sources: Version A (new PM), Version B (standard PM), or looking directly into the legislation/guidelines.

Regarding the functional suitability of the Fact Model, subjects were asked to imagine that, while carrying out their daily tasks, they faced doubts regarding what information should be taken into account within the scope of a task. In this situation, participants were questioned about their preferred information source for resolving their doubts: Version C (new FM), Version D (standard PM), or legislation/guidelines. Participants then stated their preferences on a six-point scale ranging from 1 = *definitely not* to 6 = *definitely yes*.

The instrument included empirically testable predictions to evaluate two dimensions of the SEQUAL framework [16] related to the perceived quality of representations: empirical quality (EQ) and social pragmatic quality (SPQ). In diagrammatical representations, EQ is regarded as a characteristic of the layout, which is perceived as granting it better understandability [16], whereas SPQ is the relationship between the model and the stakeholder's interpretation of the model [16].

The EQ was assessed using five questions (e.g., ''is it aesthetically attractive?''). There were five other prompts to assess the SPQ (e.g., ''is it easy to read?''). These dimensions were evaluated using a 10-item instrument on a six-point scale ranging from 1 = *strongly disagree* to 6 = *strongly agree*. Two questions were negatively phrased and reversed before the statistical analyses. The EQ scale revealed good internal consistency (5-items, $N = 32$, Cronbach's $\alpha = .82$). The SPQ scale also showed good internal consistency (5-items, $N = 32$, Cronbach's $\alpha = .84$). Participants were instructed to assess, in the exact same 10-item scale, the representations of the Process Model (Version A/B) and the Fact Model (Version C/D).

In addition to gathering demographic information (age, gender, educational level, and background), participants were requested to provide a self-assessment of their familiarity with the domain being modeled and with the DEMO standard on a six-point scale ranging from 1 = *null* to 6 = *very good*. Samples 1 and 2 evaluated representations of models related to their knowledge domains. Sample 3, composed of students, rated representations of models on urban appraisal domain (same as Sample 1).

An open question, ''Comments and suggestions for improvement'', was attached to a separate page to ensure that participants' comments were not associated with their previous answers.

2) PROCEDURE

Each study began with an introductory session aimed at informing the participants about the purpose of the research they were involved in and obtaining their informed consent. All participants voluntarily formalized their intention to participate.

Subsequently, a small training session was conducted, facilitated by an instructor certified in DEMO methodology. The purpose of this session was to ensure that the participants understood the main concepts behind DEMO representations and associated theories. The instructor presented the main concepts in DEMO and the associated theories to provide an informed and coherent interpretation of the representations being evaluated.

Following this introductory session, the participants in Sample 1 were introduced to both the standard DEMO PM representation (Version B, depicted at the top of Fig. 3) and the newly proposed representation (Version A), including the Process Diagram (bottom of Fig. 3) and Transaction Description Table (Fig. 5) [10]. For participants in Sample 2, the order of presentation of the versions was reversed to avoid potential bias related to primacy/recency effects. Sample 3 followed the same procedure as Sample 1. After the presentations were concluded, participants were provided with the aforementioned questionnaire regarding the evaluation of the PM representations.

Immediately after completing this questionnaire, another presentation session started, where the standard Object Fact Diagram from DEMO was introduced as Version D (Fig. 6, top), and the newly proposed representation (bottom of Fig. 6 and Fig. 7) as Version C [11]. The order of presentation was inverted for Sample 2 for the same reasons mentioned above. Participants were then asked to complete the questionnaire to evaluate the FM representations.

Statistical data analyses were performed using computer software (IBM SPSS Statistics, version 27 for MacOS X).

[1] https://bit.ly/QuestQualFunc

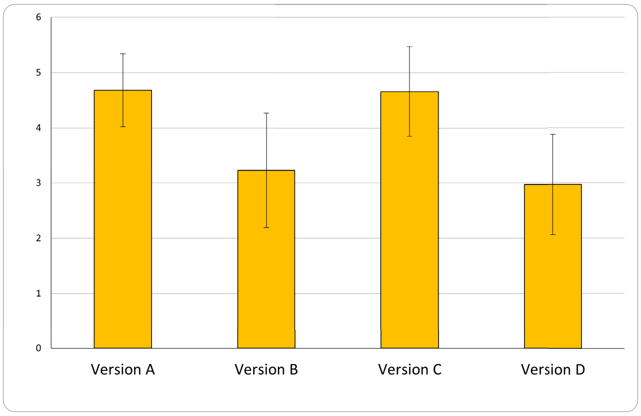


FIGURE 10. Overall evaluation of the EQ.

VI. RESULTS

This section presents the findings of our multi-sample combined analysis of user-perceived empirical quality (EQ), social pragmatic quality (SPQ), and functional suitability in representations developed under the DEMO standard. We start by examining quantitative data, and in the last section of this chapter, we analyze qualitative data.

We compared the user ratings of versions A/B and C/D. In previous individual analyses of this family of studies, we ran tests for each item using the constructs of the EQ and SPQ. Most of the data distributions were not normal; therefore, we chose to use a non-parametric tool, the Wilcoxon rank-sum test. In the new combined analysis presented in this paper, the aggregated quantitative data for the variables EQ, SPQ, and functional suitability from the three samples ($N = 32$) were normally distributed. Thus, we opted to use more robust parametric tests, and report the results in this section.

To further explore the potential variations in user perception, we ran a subgroup analysis for: sample; age group; domain knowledge; and DEMO language knowledge. Although these subgroups were relatively small (from 8 to 14 participants), the assumption of homogeneity of variance was not violated, and the assumption of normality was violated in only 4% of the cases (5 out of 120). Therefore, we decided to use only parametric tests. This choice aligns with standard procedures, as parametric tests are generally preferred when data meets underlying assumptions [42].

A. USER PERCEPTION OF EMPIRICAL QUALITY

An overall analysis of the evaluation of the EQ, considering all samples, showed that participants preferred Versions A and C (Fig. 10). This preference is still clearly visible when analyzing the samples individually (Fig. 11), even though Sample 3 does not show such a strong preference as the other groups.

To further exploit this preference, we employed a paired-sample t-test to assess user perception of EQ between the two different versions of DEMO representations across three independent samples. Both Versions A and C consistently

TABLE 3. Mean (*SD*), *t*-Statistic, and *p*-Value for user perception of EQ in Version A vs B, across samples.

	Version A		Version B					
	M	SD	M	SD	t	df	p	η^2
Sample 1	4.18	0.71	2.7	1.04	3.09	7	0.05	0.58
Sample 2	5.11	0.56	2.7	0.51	10.39	9	0.01	0.94
Sample 3	4.66	0.5	3.91	0.97	2.77	13	0.05	0.52

TABLE 4. Mean (*SD*), *t*-Statistic, and *p*-Value for user perception of EQ in Version C vs D, Across samples.

	Version C		Version D					
	M	SD	M	SD	t	df	p	η^2
Sample 1	4.48	0.69	2.8	0.68	4.13	7	0.01	0.71
Sample 2	4.7	1.1	2.34	0.61	7.55	9	0.01	0.89
Sample 3	4.73	0.66	3.51	0.92	4.16	13	0.01	0.71

received significantly higher EQ ratings than their respective counterparts (Versions B and D) in all samples.

Version A exhibited a significantly better perceived empirical quality (the mean scores, t-Statistic, p-values, and effect sizes are provided in Table 3). Similarly, Version C consistently received higher EQ ratings (Table 4).

1) ANALYSIS OF VARIANCE FOR USER PERCEPTION OF EQ ACROSS VERSIONS

A one-way between-group ANOVA was conducted to explore the differences in user perception of the EQ across the three samples for each version of the representations (A/B/C/D).

a: VERSION A

A statistically significant difference was found ($F(2, 29) = 5.75$, $p < 0.01$) and, based on omega squared, the effect size (0.23) indicated a large effect [42] (Table 5). Post-hoc Tukey's HSD tests revealed slightly significant differences in the mean EQ scores between all samples (Table 3).

b: VERSION B

A statistically significant difference was observed ($F(2, 29) = 7.6$, $p < 0.01$) with a large effect size ($\omega^2 = 0.29$) [42] (Table 5). Post-hoc comparisons using Tukey's HSD identified a significant difference in the mean EQ scores between Sample 3 and both Samples 1 and 2 (Table 3). Samples 1 and 2 did not differ from one another (Table 3).

c: VERSION C

No statistically significant difference was found in user perception of EQ across samples (Tables 4 and 5).

d: VERSION D

A statistically significant difference was detected ($F(2, 29) = 6.02$, $p < 0.01$) with a large effect size, ($\omega^2 = 0.27$) [42]

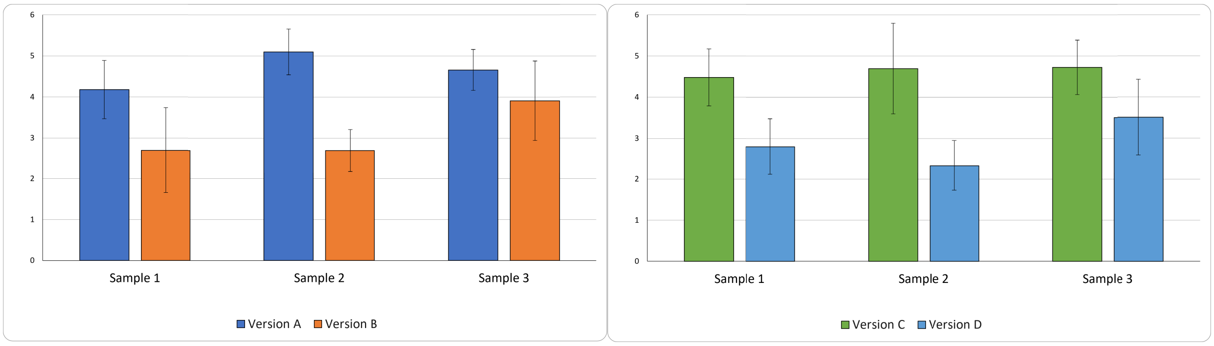


FIGURE 11. Evaluation of the EQ of Version A/B and C/D, per sample.

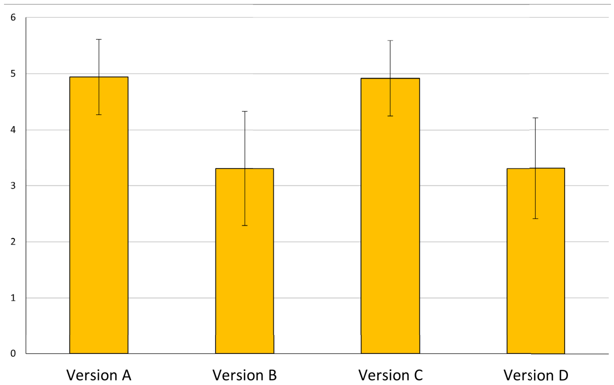


FIGURE 12. Overall evaluation of the SPQ.

(Table 5). Post-hoc Tukey's HSD tests indicated a significant difference in the mean EQ scores between Samples 2 and 3 (Table 4). Sample 1 did not differ significantly from Samples 2 and 3 (Table 4).

B. USER PERCEPTION OF SOCIAL PRAGMATIC QUALITY

Analyses of user evaluations revealed a significant preference for Versions A and C regarding the SPQ (Fig. 12). This preference was consistent across the samples (Fig. 13). Although Sample 3 also showed a trend favoring Versions A and C, this effect was less pronounced.

To better understand these differences, we employed a paired-sample t-test to evaluate the user perception of SPQ across three independent samples. Participants in all samples consistently rated Versions A and C significantly higher on the SPQ than their counterparts, Versions B and D.

Version A vs. Version B: Version A exhibited a significantly better perceived SPQ (mean scores, t-Statistic, p-values, and effect sizes are provided in Table 6).

Version C vs. Version D: Similarly, Version C consistently received higher SPQ ratings (Table 7).

1) ANALYSIS OF VARIANCE FOR USER PERCEPTION OF SPQ ACROSS VERSIONS

A one-way between-subjects ANOVA was conducted to explore potential differences in user perception of the SPQ across the three independent samples for each version of the representations (A/B/C/D).

TABLE 5. Univariate ANOVA for user perception of EQ and SPQ across versions.

		F	df	p^*	ω^2
EQ					
	Version A	5.75	2.29	0.01	0.23
	Version B	7.6	2.29	0.01	0.29
	Version C	*ns*	*ns*	*ns*	*ns*
	Version D	6.02	2.28	0.01	0.27
SPQ					
	Version A	7.02	2.29	0.01	0.27
	Version B	6.31	2.29	0.01	0.25
	Version C	*ns*	*ns*	*ns*	*ns*
	Version D	4.27	2.28	0.01	0.33

* Using a Bonferroni adjusted alpha level of 0.017
ns = Non-significant at a confidence level of 95%

TABLE 6. Mean (*SD*), *t*-Statistic, and *p*-Value for user perception of SPQ in Version A vs B, across samples.

	Version A		Version B					
	M	SD	M	SD	t	df	p	η^2
Sample 1	4.5	0.68	2.95	1.08	2.76	7	0.05	0.52
Sample 2	5.46	0.42	2.72	0.65	10.9	9	0.01	0.94
Sample 3	4.81	0.58	3.93	0.9	3.59	13	0.01	0.65

TABLE 7. Mean (*SD*), *t*-Statistic, and *p*-Value for user perception of SPQ in Version C vs D, across samples.

	Version C		Version D					
	M	SD	M	SD	t	df	p	η^2
Sample 1	4.58	0.74	3.03	0.65	3.69	7	0.01	0.66
Sample 2	4.88	1.24	2.6	0.71	6.75	9	0.01	0.87
Sample 3	4.91	0.65	3.87	0.82	3.73	13	0.01	0.67

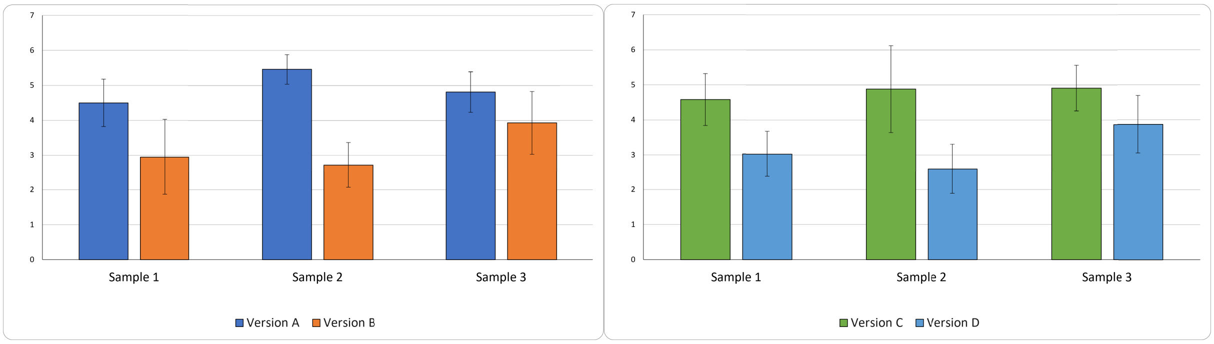


FIGURE 13. **Evaluation of the SPQ of Version A/B and C/D, per sample.**

a: VERSION A

A statistically significant difference was observed in user ratings of SPQ ($F(2, 29) = 7.02$, $p < 0.01$). Furthermore, based on the omega squared, the effect size indicated a large effect ($\omega^2 = 0.27$) [42] (Table 5). Post-hoc Tukey HSD tests revealed a significant difference in the mean SPQ scores between Sample 2 and, Sample 1 and 3 (Table 6). Samples 1 and 3 exhibited no significant differences (Table 6).

b: VERSION B

A statistically significant difference was detected in the SPQ ratings ($F(2, 29) = 6.31$, $p < 0.01$) with a large effect size, ($\omega^2 = 0.25$) [42] (Table 5). Post-hoc Tukey HSD comparisons identified a significant difference in the mean SPQ scores between Sample 3 and both Sample 1 and 2 (Table 6). Samples 1 and 2 showed no significant differences (Table 6).

c: VERSION C

No statistically significant difference was found in user perception of SPQ across samples (Tables 5 and 7).

d: VERSION D

A statistically significant difference was identified in the SPQ ratings ($F(2, 28) = 4.27$, $p < 0.01$) with a large effect size, ($\omega^2 = 0.33$) [42] (Table 5). Post-hoc Tukey HSD tests indicated a significant difference in the mean SPQ scores between Sample 3 and both Samples 1 and 2 (Table 7). Samples 1 and 2 did not differ significantly from one another (Table 7).

C. VARIANCE OF USER PERCEPTION OF EQ AND SPQ, ACROSS VERSIONS AND SAMPLES

A two-way between-group MANOVA was conducted to investigate the effects of sample (1, 2, and 3) and version (A/B or C/D) on user perception of the EQ and SPQ. Preliminary assumption testing was performed to check for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity, with no serious violations noted.

a: VERSION A/B

There was a statistically significant interaction effect between the sample and version ($F(8, 52) = 3.78$, $p < 0.01$; Wilks' $\lambda = 0.4$). Univariate ANOVA with Bonferroni correction revealed significant main effects of version for both the EQ and SPQ (all p-values < 0.01, see Table 5). Sample 2 consistently rated Version A higher than Samples 1 and 3 for both measures (Tables 3 and 6). Conversely, Sample 3 rated Version B higher than Samples 1 and 2 on both measures (Tables 3 and 6).

b: VERSION C/D

A significant interaction effect between sample and version was observed ($F(8, 50) = 2.82$, $p < 0.05$; Wilks' $\lambda = 0.54$). However, only Version D showed a significant main effect of version for both the EQ and SPQ after Bonferroni correction (all p-values $< .01$, see Table 5). Similar to Version B, Sample 3 rated Version D higher on both measures than Samples 1 and 2 (Tables 4 and 7).

D. VARIANCE OF USER PERCEPTION OF EQ AND SPQ, BY AGE GROUP

A one-way between-group ANOVA was conducted to explore the effects of age on user perception of the EQ and SPQ for each version (A/B/C/D). The participants were divided into three groups: Group 1: 25 years old or less ($N = 11$), Group 2: 26 to 38 years old ($N = 11$, except on SPQ of Version C where $N = 10$), and Group 3: 39 years old and above ($N = 10$).

In Version B, we found a significant effect of age group for both EQ ($F(2, 29) = 9.08$, $p < 0.01$, $\omega^2 = 0.34$) and SPQ ($F(2, 29) = 7.12$, $p < 0.01$, $\omega^2 = 0.28$). Post-hoc Tukey's HSD tests revealed that younger participants rated Version B significantly higher ($N_{EQ} = 4.11$, $SD_{EQ} = 0.99$; $N_{SPQ} = 4.09$, $SD_{SPQ} = 0.94$) than both the 26-38 age group ($M_{EQ} = 2.71$, $SD_{EQ} = 0.48$; $M_{SPQ} = 2.76$, $SD_{SPQ} = 0.69$) and older participants ($M_{EQ} = 2.84$, $SD_{EQ} = 0.97$; $M_{SPQ} = 3.04$, $SD_{SPQ} = 0.96$).

Another significant effect was observed for Version D, both in the evaluation of EQ ($F(2, 29) = 10.09$, $p < 0.001$, $\omega^2 = 0.36$) and SPQ ($F(2, 29) = 11.46$, $p < 0.001$, $\omega^2 = 0.4$). Post-hoc Tukey HSD tests indicated that younger participants rated Version D significantly higher ($M_{EQ} = 3.73$, $SD_{EQ} = 0.89$; $M_{SPQ} = 4.09$, $SD_{SPQ} = 0.78$) than both the 26-38 age group ($M_{EQ} = 2.36$, $SD_{EQ} = 0.54$; $M_{SPQ} = 2.76$,

$SD_{SPQ} = 0.79$) and the 39+ age group ($M_{EQ} = 2.80$, $SD_{EQ} = 0.71$; $M_{SPQ} = 2.9$, $SD_{SPQ} = 0.5$).

There were no statistically significant differences in user perceptions of the EQ and SPQ for Versions A and C across the age groups.

E. VARIANCE OF USER PERCEPTION BY SELF-REPORTED KNOWLEDGE

We controlled for self-reported DEMO and domain knowledge to better understand how participants rated the EQ, SPQ, and functional suitability.

1) ANALYSIS OF VARIANCE FOR USER PERCEPTION BY SELF-REPORTED DEMO KNOWLEDGE

A one-way between-group analysis of variance was conducted to explore the differences between self-reported DEMO knowledge and the evaluation of EQ, SPQ, and functional suitability of Versions A/B/C/D. Subjects were divided into three groups according to their self-reported knowledge of the DEMO language: Group 1: Low Knowledge ($<= 2$ points, $N = 11$), Group 2: Medium Knowledge ($3 - 4$ points, $N = 12$), and Group 3: High Knowledge (5+ points, $N = 9$).

There were no statistically significant differences in the EQ or SPQ in Versions A, C, and D, across all three DEMO knowledge groups.

A statistically significant difference was observed in the evaluation of both the EQ ($F(2, 29) = 4.42$, $p < 0.05$, $\omega^2 = 0.18$) and SPQ ($F(2, 29) = 5.15$, $p < 0.05$, $\omega^2 = 0.21$), in Version B. Post-hoc comparisons using Tukey's HSD tests indicated that participants with lower self-reported DEMO knowledge ($M_{EQ} = 2.58$, $SD_{EQ} = 0.71$; $M_{SPQ} = 2.64$, $SD_{SPQ} = 0.54$) rated Version B significantly lower than participants with middle ($M_{EQ} = 3.4$, $SD_{EQ} = 1.11$; $M_{SPQ} = 3.47$, $SD_{SPQ} = 1.09$) and higher knowledge ($M_{EQ} = 3.8$, $SD_{EQ} = 0.96$; $M_{SPQ} = 3.91$, $SD_{SPQ} = 1$).

There was one statistically significant difference in the evaluation of functional suitability using Version D between the groups with the lowest and highest self-reported DEMO knowledge ($F(2, 29) = 5.2$, $p < .05$). The effect size, calculated using the omega squared, was small ($\omega^2 = 0.3$). Post-hoc comparisons using Tukey's HSD test also indicated that the functional suitability mean score for the group with lower knowledge ($M = 2.55$, $SD = .69$) was significantly lower than that of the group with the highest DEMO knowledge ($M = 4.11$, $SD = 1.05$).

There were no statistically significant differences in the evaluation of functional suitability in Versions A/B/C across the three DEMO knowledge groups.

2) ANALYSIS OF VARIANCE FOR USER PERCEPTION, BY SELF-REPORTED DOMAIN KNOWLEDGE

A one-way between-group ANOVA was conducted to explore the effects of the differences in self-reported domain knowledge on user perception of the EQ and SPQ for Version A/B/C/D. The subjects were divided into three groups according to their self-reported domain knowledge (Group 1: $<= 3$, $N = 15$; Group 2: 4, $N = 9$; Group 3: 5+, $N = 8$). No statistically significant differences were found across the domain knowledge groups regarding user perceptions of the EQ and SPQ.

We also ran an ANOVA to explore the differences between the level of self-reported domain knowledge and the evaluation of the functional suitability of Versions A/B/C/D. There was a statistically significant difference in the evaluation of Version D between the two groups with self-reported lower domain knowledge ($F(2, 29) = 4.43$, $p < .05$, $M_{Group1} = 3.87$, $SD_{Group1} = 1.06$, $M_{Group2} = 2.67$, $SD_{Group2} = 1$). No effect size was detected using omega squared. There were no statistically significant differences in the functional suitability of Versions A/B/C across the knowledge groups.

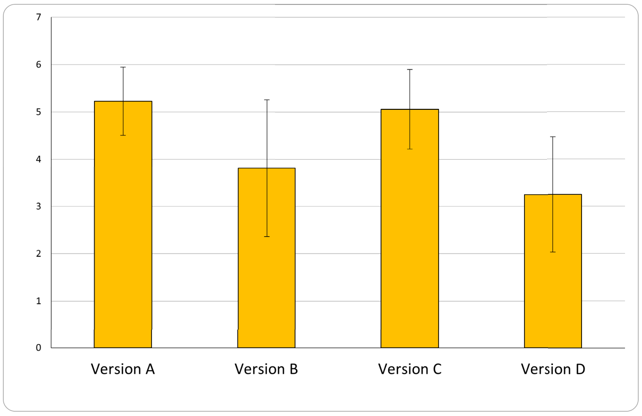


FIGURE 14. **Overall evaluation of the functional suitability.**

F. USER PERCEPTION OF FUNCTIONAL SUITABILITY

An overall observation of the ratings given for functional suitability shows that participants considered Versions A and C to be more functional (Fig. 14). Looking at the evaluation per sample in Fig. 15, we can see that Versions A and C are preferred.

A paired-sample t-test was conducted to compare user perception of functional suitability between Versions A/B and C/D. Functional suitability was rated significantly higher for Version A than for Version B ($M_A = 5.23$, $SD_A = 0.72$; $M_B = 3.81$, $SD_B = 1.45$; $t(30) = 4.68$, $p < 0.001$). The eta square statistic ($\eta^2 = 0.42$) indicated a large effect size [42]. Version C also received significantly higher functional suitability ratings than Version D ($M_C = 5.06$, $SD_C = 0.84$; $M_D = 3.25$, $SD_D = 1.22$; $t(30) = 7.71$, $p < 0.001$). The effect size ($\eta^2 = 0.67$) was again large [42].

To further evaluate the perception of functional suitability, participants compared Versions A/B and C/D across three criteria: ease of understanding task sequence/fact relationships, ease of viewing and understanding tasks/facts, and suitability for supporting everyday tasks. For all three criteria, the participants rated Versions A and C significantly higher than Versions B and D (Fig. 16 and 17).

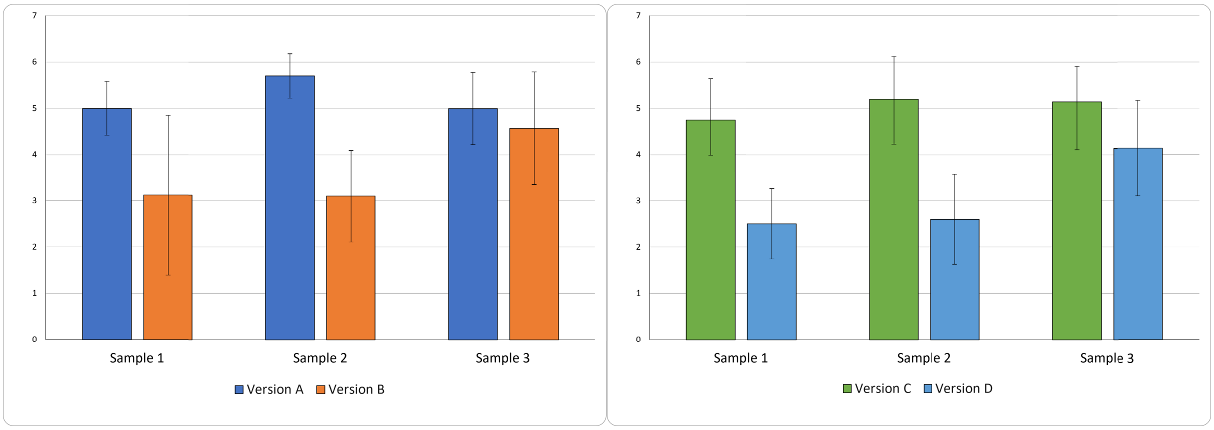


FIGURE 15. Evaluation of the functional suitability, per sample.

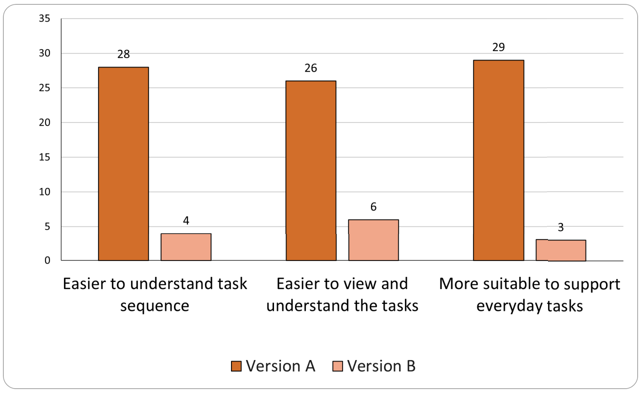


FIGURE 16. Functional suitability of Version A/B.

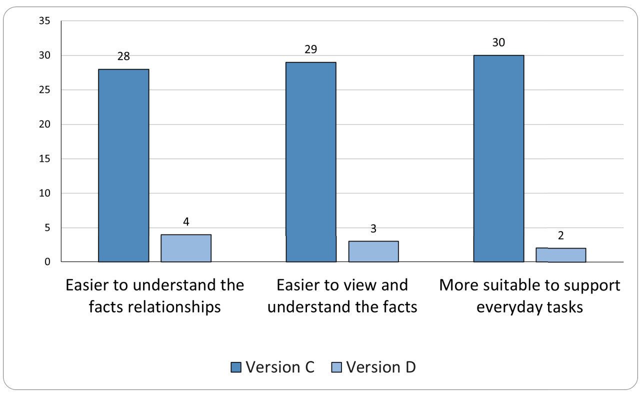


FIGURE 17. Functional suitability of Version C/D.

1) USER PERCEPTION OF FUNCTIONAL SUITABILITY BY SAMPLE

A paired-sample t-test was conducted to compare the user perception of functional suitability between Versions A/B and Versions C/D within each sample. Versions A and C were systematically rated higher than Versions B and D (see Tables 8 and 9).

TABLE 8. Mean (*SD*), *t*-Statistic, and *p*-Value for functional suitability in Version A/B, across samples.

	Version A		Version B					
	M	*SD*	*M*	*SD*	*t*	*df*	*p*	η^2
Sample 1	5	0.58	3.29	1.8	*ns*	*ns*	*ns*	*ns*
Sample 2	5.7	0.48	3.1	0.99	6.5	9	0.001	0.86
Sample 3	5	0.78	4.57	1.22	*ns*	*ns*	*ns*	*ns*

ns = Non-significant at a confidence level of 95%

TABLE 9. Mean (*SD*), *t*-Statistic, and *p*-Value for functional suitability in Version C/D, across samples.

	Version C		Version D					
	M	*SD*	*M*	*SD*	*t*	*df*	*p*	η^2
Sample 1	4.75	0.89	2.5	0.76	*4.28*	7	0.01	0.72
Sample 2	5.2	0.92	2.6	0.97	8.51	9	0.001	0.91
Sample 3	5.14	0.77	4.14	1.03	3.61	13	0.001	0.65

2) USER PREFERENCE FOR INFORMATION SOURCES IN TASK/FACTS CLARIFICATION

In Fig. 18, one can see the overall preference of the participants for the preferred information source.

A paired-sample t-test was conducted to compare user preferences for different information sources when clarifying questions regarding specific tasks.

a: VERSION A VS. VERSION B

Participants showed a significantly stronger preference for Version A ($M = 5.06$, $SD = 0.93$) than for Version B ($M = 3.19$, $SD = 1.40$) for task-related questions ($t(30) = 5.66$, $p < 0.001$, $\eta^2 = 0.52$).

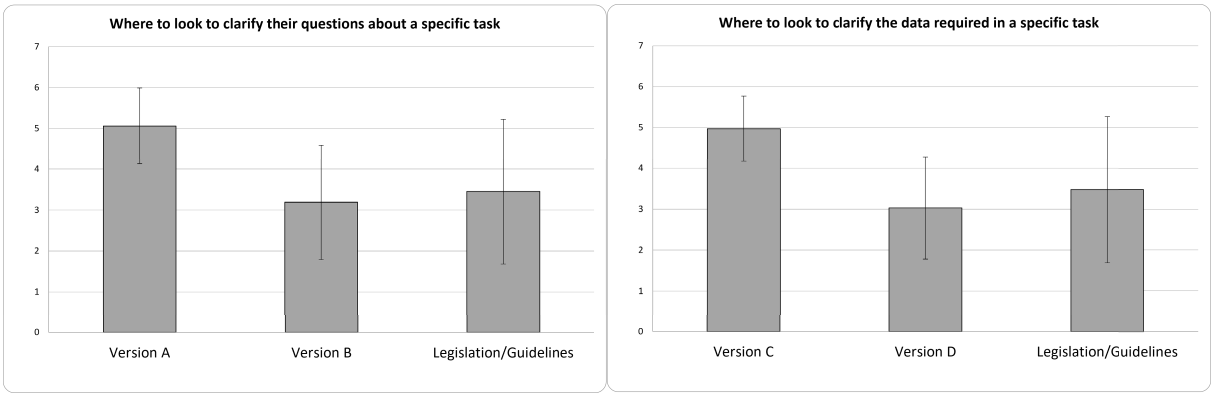


FIGURE 18. Preferred information source.

b: VERSION A VS. LEGISLATION/GUIDELINES

Version A was also significantly preferred over legislation/guidelines ($M = 3.45$, $SD = 1.77$) when clarifying task-related questions ($t(30) = 4.29$, $p < 0.001$, $\eta^2 = 0.38$).

c: VERSION C VS. VERSION D

Participants preferred Version C ($M = 4.97$, $SD = 0.8$) over Version D ($M = 3.03$, $SD = 1.25$) to clarify questions about data requirements within tasks ($t(30) = 7.06$, $p < 0.001$, $\eta^2 = 0.62$).

d: VERSION C VS. LEGISLATION/GUIDELINES

Version C again received a significantly higher preference than the legislation/guidelines ($M = 3.48$, $SD = 1.79$) for clarifying data requirements within tasks ($t(30) = 3.91$, $p < 0.001$, $\eta^2 = 0.34$).

G. QUALITATIVE DATA ANALYSIS

In a debriefing session, after completing the questionnaires, participants expressed comments and opinions on what could be improved in the representations. This feedback was reviewed in light of qualitative research methods, namely, narrative analysis.

In Sample 1, the features that gathered the most consensus were as follows: inclusion in the Process Diagram (Version A), the time restrictions of each task; and present, associated with the tasks, the organizational role responsible for executing the task in some sort of overlay. Participants from Sample 1 revealed that they would like to have included the exception conditions of the task in the task shape (e.g., a collapsible box) and the procedures that apply in each case (e.g., a task not finishing within the expected timeframe or other conditions such as the delivery of certain mandatory documents not being fulfilled). This information is already available in the transaction description table, but they considered it to be helpful to have it represented in the diagrams.

The participants in Sample 1 emphasized that they would need to update the representations frequently because of the new rules and procedures that are continuously created and updated. Moreover, the participants reinforced the need to include informal descriptions in diagrams.

In Samples 1 and 2, participants mentioned that Version B was difficult to grasp and had extensive line clutter.

VII. DISCUSSION

In this section, we explore and discuss key findings related to our results and hypotheses.

A. OVERALL PERCEIVED QUALITY AND FUNCTIONAL SUITABILITY

The results from our combined analyses indicate that the newly proposed representations of DEMO's Process and Fact Models express information in a concise, clear, effective, and useful manner, supporting H_1. The results of the evaof the functional suitability of versions A and C, as well as their EQ and SPQ, show a clear advantage over versions B and D (Fig. 10, 12, and 14). Results of EQ and SPQ ratings are aligned; therefore, we consider that the overall perceived quality of the new representations is higher, as detailed in the following sections. Together with the explicit results regarding functional suitability, the aggregated data contributed to validating H_1.

Our findings suggest that diagrammatical representations with concise information, complemented by semantically rich descriptive tables containing comprehensive and linked information, seem to be perceived as having higher quality and functional suitability. This is also supported by previous work and the Cognitive Theory of Multimedia Learning, which hypothesizes that material intended to be received, understood, and retained by its users should be presented using both words and visuals [26].

Considering the limitations of working memory and the concept of element interactivity pointed out by the Cognitive Load Theory [28], the extraneous cognitive load should be decreased as far as possible [27]. The newly proposed representations are clear and concise, which is essential to reduce the cognitive effort required to process them effectively. Therefore, these representations seem to overcome the toll of the standard representations being too complex and not guaranteeing model comprehensibility for stakeholders, a concern reported by several authors (e.g., [2], [3], [7]). Thus, Versions A and C seem to have achieved fitness for use by their audience [3] and can aid in more effective enterprise modeling for more successful information system planning and implementation.

B. EMPIRICAL QUALITY

The combined analysis found that Versions A and C were consistently rated as having a higher EQ than Versions B and D (Fig. 10. This suggests that the newly proposed representations are regarded better through a different arrangement or layout, confirming that they have a higher EQ (H_2) [16].

Several design aspects of Versions A and C likely contributed to this perception (e.g., color, solidity, aesthetics, and orientation). Incorporating symbols and notations that facilitate the understanding of the modeled system seems to be a clear advantage of the new representations. The participants' feedback also highlighted the notation's aesthetics that facilitated task identification.

Version A used color to differentiate between original, informational, and documental tasks, potentially aiding information identification and comprehension, as supported by previous work [29]. Even though the literature reports conflicted results regarding the use of color in process models [4], our results show that it helped users' understanding.

In the debriefing session, participants mentioned that they would like to have information on the exception conditions and the procedure to deal with them, included or accessible in its shape through a collapsible box (for example, similar to our innovation of the collapsible box in the FM's concepts). This supports previous findings that representations should include dynamic features [19] and informal descriptions [16], [19].

Retinal variables (e.g., shape, size, and color) were used by Version C to accentuate information, potentially paving the way for a higher perceived EQ, confirming findings from previous studies [29]. Additionally, its dynamic nature, allowing users to hide/show details and rearrange elements, might have been interpreted as a sign of better organization and, thus, higher EQ.

Unlike the static and cluttered feel of Version D, the higher aesthetic appeal of Version C was probably due to the collapsible boxes. As claimed in previous work [11], this design feature leads to more aesthetically pleasant information visualization and allows users to focus on key aspects. Because of its complexity, the current OFD (in GOSL) can be a burden for stakeholders who are untrained in the DEMO standard [11]. The findings of our family of studies indicate that the newly proposed representations overcome this limitation.

According to the literature, increased familiarity with a diagram can positively impact the assessment of its aesthetics [16]. Our findings partially support this hypothesis. The analysis revealed consistency in evaluating Versions A and C across the samples (Tables 3 and 4), suggesting that the cognitive effectiveness of these representations transcends DEMO knowledge. However, the ratings for Versions B and D differed significantly (Fig. 11). While Samples 1 and 2 (with no DEMO experience) showed similar evaluations for these versions, Sample 3 (DEMO-trained) rated them significantly higher. This suggests that DEMO knowledge enhances user perception and positively influences EQ evaluation. Nevertheless, even in Sample 3, newer representations were preferred. These results did not reveal a positive association between, on the one hand, participants' self-reported DEMO knowledge and, on the other hand, the evaluation of the EQ in Versions B and D. Thus, our initial expectation was not fulfilled (H_6).

Controlling for self-reported domain knowledge, we found no significant differences in the evaluation of EQ, thus not supporting H_5, regarding the EQ scale. This seems to indicate that knowledge about modeled tasks and facts does not necessarily interfere with the perception of EQ. It implies that the notation used is domain-neutral, which is a required property for a generic language.

When controlling for age, we found that younger participants rated the EQ of Versions B and D significantly higher. This might be because most younger participants belonged to Sample 3 (DEMO-trained); therefore, the representations on Version B and D were already familiar to them. Nevertheless, only Version B showed a significant difference when controlling for self-reported DEMO knowledge in both the EQ and SPQ. Participants with higher self-reported DEMO knowledge (groups with medium and high self-reported knowledge; that is, level 3+) evaluated the EQ of Version B more positively. This suggests that the struggle to understand the traditional Process Model representation is lower for those with formal language knowledge.

Symbols are an essential aspect of EQ [16]. MU-theory [12] explains and clarifies the notion of models and representations. However, DEMO symbols are not currently grounded in theory or empirical studies. As we can infer from our family of studies and also pointed out by other authors [7], [8], [18], [19], [20], the traditional DEMO Process and Fact Model representations are inadequate communication tools, somewhat cluttered, and difficult for laypeople to understand. A more user-friendly representation of models, with higher perceived quality and functional suitability is required to elevate the general adoption rate of DEMO standard by industry. The results we present constitute relevant steps in this direction.

C. SOCIAL PRAGMATIC QUALITY

Participants in our family of studies consistently rated Versions A and C as having a higher SPQ than Versions B and D (Fig. 12), thereby confirming H_3. This suggests that newer representations were perceived as easier to understand, probably because of their features of effectively highlighting relevant content, improved structure, intelligibility, readability, and comprehensiveness, thus having a higher SPQ [16]. That is, Versions A and C are better understood by their intended audience [16], [29].

In the analyses of data per sample (Table 6 and 7), we found that Sample 3 (DEMO-trained) rated Versions B and D significantly higher than the other samples with no DEMO experience (Fig. 13). This finding aligns with the literature, as it has been reported that familiarity with a modeling language can improve its understandability [16].

Despite this, even among Sample 3 (DEMO-trained), Versions A and C were still the preferred choices, not supporting our initial prediction (H_6) regarding the SPQ. This might be because the subjects of Sample 3 were not sufficiently experienced in the DEMO standard and, hence, preferred a mixed format for representing the Process and Fact Models over a purely formal one. Future studies should explore the relationship between proficiency in formal languages and perceived intelligibility of diagrammatic representations.

Controlling for self-reported DEMO knowledge, we found only one statistically significant difference in SPQ. In the evaluation of Version B, namely in the groups with higher DEMO knowledge (self-reported level of 3+), the SPQ was evaluated more positively. Considering this finding, we can infer that participants with lower DEMO knowledge may have felt challenged to understand the standard Process Model representations.

According to our findings, self-reported DEMO knowledge did not significantly influence SPQ evaluation for Versions A, C, or D. This may be due to the small number of participants with high self-reported DEMO knowledge ($N = 9$) in our sample, which did not provide enough data for a more generalizable analysis. Further research is needed to explore the relationship between expertise in the DEMO standard and SPQ evaluation.

No statistically significant differences were found regarding the effects of self-reported domain knowledge on SPQ perception. This suggests that domain knowledge itself may not influence SPQ perceptions.

Krogstie [29] reported that a mix of formal and informal statements can be more comprehensive for subjects without modeling training than purely formal models, even if the information present is redundant. Version C of the newly proposed representations incorporates this mix of formal and informal statements and, therefore, potentially contributes to participants' preferences in Samples 1 and 2 (no DEMO experience). Conversely, subjects trained in structured visual notations tend to prefer formal to informal models. However, the data from our family of studies did not confirm this, as Sample 3 was trained in the DEMO standard, but attributed a higher evaluation to the newly proposed representations.

No significant differences were found when self-reported domain knowledge was controlled for. Our results do not support H_5, in terms of the SPQ. Therefore, being knowledgeable about modeled concepts does not seem to affect the evaluation of the SPQ.

When controlling for age, we found that younger participants (mostly belonging to Sample 3) rated the SPQ of Versions B and D significantly higher than the remaining participants. This is in line with the previously reported preference for traditional DEMO representations within this group, which had undergone DEMO training. No other significant age-related variations were observed in Versions A and C.

D. FUNCTIONAL SUITABILITY

Versions A and C were perceived as more suitable, that is, those that more satisfactorily addressed stated and implied needs, supporting H_4 (Fig. 14) [21]. Version D was perceived to be less functional.

Going deeper into the data (Tables 8 and 9), we found that all three samples considered Versions A and C to be more functional (see Fig. 15). Furthermore, we found that the sample that best classified the functional suitability of the standard DEMO diagrams (Versions B and D) was Sample 3 (DEMO-trained). However, these results do not confirm H_6, as participants with higher self-reported DEMO knowledge still preferred the newly proposed DEMO representations. These results come in line with previous studies focused on label types in process modeling, which also could not find a statistically significant interaction effect on the relationship between familiarity with the notation and the evaluation of perceived usefulness [31].

Our results were inconclusive for H_5, as they did not reach statistical significance. Previous studies focused on label types in process modeling, also could not find a significant interaction effect on the relationship between domain knowledge and its perceived usefulness [31]. Further evaluations with larger samples are needed to reach statistically significant interaction effects.

Although not statistically significant, the group with higher knowledge (Group 3, knowledge level ≥ 5) also assessed the functional suitability of Version D as lower than that of Group 1. This might be explained by the fact that Group 1 had less knowledge of the modeled concepts and, therefore, did not fully realize the difficulties of interpreting the representation.

A large majority of the subjects directly chose the new representations as: the easiest to enable understanding of task sequence and fact relationships; the easiest to visualize and understand tasks/facts; and the most suitable to support the execution of their daily tasks (Fig. 16 and 17).

According to Krippendorff, the worth of any design is judged in accordance with the perceptions and meanings that users attribute to it [9]. Our data revealed that when needing to clarify a question, a large majority of the participants preferred to look it up in the new representations (Versions A and C) rather than in the traditional representations (Versions B and D) or even in the legislation/guidelines (Fig. 18), supporting H_7.

The results confirm that participants perceived new representations as meeting stated and implied needs [21], thus having higher functional suitability.

VIII. THREATS TO VALIDITY AND LIMITATIONS

While this family of studies offers valuable insights into user perception of DEMO representations, threats to validity and limitations were considered. These factors were considered both in the research design to strengthen the validity of our results and, in this section, to provide inputs and directions on how future research can address current limitations. Threats to validity within the scope of empirical Software Engineering are reported in [43]. This is closest to the research presented in this paper, as models and representations are an essential part of model-driven software engineering/information system development. In quantitative research, four main types of threats to validity were considered: conclusion, internal, construct, and external validity [43].

Regarding conclusion validity, these threats could apply to our family of studies: (1) low statistical power; (2) violated assumptions of statistical tests; (3) fishing and error rate; (4) reliability of measures, and (5) random heterogeneity of subjects.

With respect to (1), our sample size was relatively small ($N = 32$) due to the resource constraints to recruit more participants. Nevertheless, we only reported statistically significant results of over 95% confidence level and the tests we used (t-test, ANOVA, and MANOVA) have high statistical power. In (2), there are no major violated assumptions, as reported in Section VI. The data distribution in our dataset was normal, and the tests used were robust. In the case of (3), we could argue that we expected the newly proposed representations to achieve a higher evaluation. However, the manner in which we designed and conducted our studies, as described in Section V, aimed to ensure independence and escape from common bias. As this is a comparative study, we looked at the data from several perspectives, as reported in Section VI. As a reference, we used the usual 95% confidence level; however, some results were significant at the 99% confidence level. In some analyses, we used Bonferroni correction, as indicated in the literature [42].

Regarding the reliability of measures (4), the validity of an experiment is highly dependent on them. The basic principle is that, when measuring a phenomenon twice, the outcome will be the same. For example, lines of code are more reliable than function points, because they do not involve human judgment. In other words, objective measures that can be repeated with the same outcome are more reliable than subjective measures. In a subjective measure, the person making the measurement contributes by making some sort of judgment, and the measurement can be different if the object is re-measured. Examples of subjective measures include personnel skills and usability [43]. Owing to the nature of our evaluation, all measures were subjective, which could imply a high risk of threat to validity. Several questions were associated with each measure to ensure reliability, some of which were negatively phrased. Regarding the outcome in our family of studies, we measured the same phenomenon three times, and the results were quite similar. Therefore, we concluded that this is not a significant threat to our study.

With respect to (5), we had a larger degree of homogeneity within samples rather than heterogeneity, which could hinder external validity; however, by having a relevant and substantial heterogeneity across samples, we mitigated this effect. Nevertheless, further research is required to increase the reliability of the results and external validity, as our family of studies relied on self-reported data collected through purposeful sampling from three user groups (public officers, healthcare professionals, and DEMO students, all living in Portugal). This limits the generalizability of our findings to other demographics and domains. Future research should involve larger and more diverse samples to explore potential variations in user perceptions across different backgrounds and geographies.

With respect to internal validity, we recognize as possible single-group threats: (1) history, (2) maturation, (3) instrumentation, (4) selection, (5) mortality, and (6) ambiguity regarding the direction of causal influence [43].

History threat (1) did not apply to our study, as we had a single intervention with each group of participants. Regarding (2), the subjects in our study might have reacted differently as time passed, as the session took 60-90 minutes. The evaluation session included seven steps: presentation of the main DEMO concepts, presentation of the two versions of the Process Model, filling in the questionnaire related to the Process Model, presentation of the Fact Model, filling in the questionnaire related to the Fact Model, filling in the demographic data, and a debriefing session. To mitigate the fact that some subjects might have been negatively (tired, bored, or worn out) or positively (learning) affected, we alternated between the presentation of the models and the time to fill in the questionnaire. With this design, we hoped to make the sessions more dynamic.

Concerning (3), the questionnaire used in this family of studies was presented in [19]. It was designed primarily based on the SEQUAL framework, which has been extensively tested [2], [16], [30]. Selection effect (4) stresses the possible effect of letting volunteers participate in an experiment, as it may influence the results. Volunteers are generally more motivated and suited to new tasks than the entire population [43]. As we used purposeful sampling, the generalizability of the findings was limited to the conclusion validity threats, as acknowledged above. Mortality (5) and ambiguity about the direction of causal influence (6) do not apply to our family of studies, as we had neither participants who dropped out nor tried to establish causal relationships among variables.

As our study involved a single interaction with each group of participants, social threats to internal validity [43] are not applicable.

In terms of construct validity, some relate to the design of the experiment, whereas others relate to social factors [43].

Regarding design threats, our constructs were well defined; thus, the design threat of inadequate preoperational explication of constructs does not apply. Our family of studies included different variables (e.g., empirical quality, functional suitability, DEMO knowledge). We used mixed methods to collect data, and the study took place in only one session per group. Therefore, the design threats of mono-operation bias, mono-method bias, and interaction do not apply.

Relevant design threats could be confounding constructs and levels of constructs. In some relations, it is not primarily the presence or absence of a construct, but the level of the construct, which is important to the outcome. The effect of the presence of a construct is confounded by the construct-level effect. For example, the presence or absence of prior knowledge in a programming language may not explain the causes of an experiment, but the difference may depend on whether the subjects have 1, 3, or 5 years of experience in the current language [43].

As already pointed out in the discussion of the results on SPQ regarding H_6, the fact that the subjects of sample 3 were at the student level and thus had less than a year of experience with the DEMO language might have led to H_6 not being supported. The fact that there was a significant (positive) difference in the evaluation of SPQ of versions B and D, while controlling for the variable of experience in DEMO, might imply that if subjects have several years of experience in the DEMO language, they might consider versions B and D to have more SPQ than versions A and C, and eventually support H_6. Therefore, future studies should recruit participants more experienced with DEMO. Nevertheless, the fact that we assessed the DEMO experience on a Likert-type scale probably mitigated this threat.

In terms of social factors, we identified possible threats: hypothesis guessing and experimenter expectations. We tried to avoid the hypothesis of guessing a threat by presenting both new and traditional representations in an independent way and, in one study, in a reversed order. Experimenters can bias the results of a study, both consciously and unconsciously, based on their expectations. This threat can be reduced by involving different subjects with no or different expectations from the experiment. For example, questions can be raised in different ways to provide the same answers [43]. To counter this threat in our family of studies, the questions were designed to minimize it, as already described in the measure of reliability threat (the same variable was questioned in different ways). However, these studies did not control for the potential bias introduced by the presenter during the initial briefings. Future studies should use multiple presenters to mitigate this potential effect.

Owing to a lack of time and resources, the evaluation of functional suitability did not include its subcomponents of completeness, correctness, and appropriateness. Future studies could, therefore, control for these variables.

IX. CONCLUSION AND FUTURE WORK

Our research, in the form of a family of studies, comparatively evaluated representations of the Process and Fact Models of the DEMO standard. We compared the newly proposed representations (Versions A and C), which arose from practical needs in a large-scale municipality modeling project, to the current standard diagrams (B and D). This study addresses the combined analysis of the pooled data from three previous studies. This analysis strongly supports the claim that the newly proposed representations are more accessible and easier to grasp, both by domain experts and those familiar with the DEMO standard. Overall, the family of studies presented herein highlights the importance of using user-centered design principles to improve cognitive effectiveness in the representations of DEMO models.

It should be noted that the representations in Versions A and C were not semantically equivalent to those in their counterparts (B and D). Versions A and C incorporated elements (from other DEMO models) pointed out as required/useful by the stakeholders of the modeling project, which were not originally part of DEMO's PM and FM, whereas information considered irrelevant to the stakeholders was removed. Therefore, rather than comparing the representations based on strict semantic equivalence, this family of studies aimed to evaluate user perception of the quality and functional suitability of alternative representations, where the newly proposed ones try to bring more semantic richness, on one hand, while also aiming, on the other hand, for simplicity. It could be argued that one should not compare representations where a particular alternative has more information and can naturally and inherently be superior to the other. However, standard DEMO representations have been proposed as a convention, and no theoretical or usability studies have been conducted on them [1]. Thus, to the best of our knowledge, ours is the first formal evaluation of user perception of the quality and functional suitability of DEMO representations. And it has a large-scale real-world practical project supporting the design of the newly proposed representations. Therefore, this comparison appears useful and fair.

This family of studies has contributed to the body of knowledge in several ways. First, we deepened and consolidated previous research examining user perception of empirical quality (EQ), social pragmatic quality (SPQ), and the functional suitability of standard and newly proposed representations of DEMO's Process and Fact Models. This strengthens the reliability of previous findings. Second, we expanded the knowledge base on the barely unexplored area of user experience with DEMO models. Third, we hope that this family of studies will drive further research on user-centric approaches, ultimately leading to enhanced DEMO representations that are easier to understand and apply. Fourth, our work stresses the importance of having diagrammatic representations that are dynamic, subject to frequent modifications, and must be easily changed. Ultimately, our family of studies provides valuable insights into stakeholders' preferences. This information can be used to enhance the Process and Fact Model representations, which can potentially increase their quality and encourage broader adoption of the DEMO standard by industry.

As a general insight, diagrammatical representations with simple and essential data complemented by descriptive tables with additional information are perceived as having higher quality and functional suitability. Further research should explore this complementary strategy to enable a better user experience in the field of modeling.

Representations of DEMO models must improve their qualities to increase their adoption by industry. Future research should focus on developing more accessible and inclusive representations, considering the insights from the Cognitive Load Theory and the Cognitive Theory of Multimedia Learning. It would benefit the full range of stakeholders, further simplifying and improving the visual notation and language used in DEMO representations. Enhancing user comprehension and cognitive effectiveness, as well as investigating the impact of different representations on task performance and user learning outcomes, will provide further understanding of their effectiveness. According to our findings, the newly proposed representations of the Process and Fact Models hold promise, even for those experienced in DEMO.

We will now focus on relevant practical implications of the newly proposed DEMO representations. Their user-centric design places emphasis on understandability and practical usability, making them accessible and functional for stakeholders.

On our family of studies, there were no statistically relevant differences regarding the evaluation of the newly proposed representations by different domains (urban appraisal and healthcare). This seems to imply that this notation has a neutral effect domain-wise, which means it should be applicable to most, if not all, organizational settings.

The new representations provide increased traceability. For example, they allow users to easily track the tasks that create or update all facts/information within the model. This facilitates verification and auditing processes.

The novelty of the descriptive tables, which provide detailed and structured information to support the interpretation of the diagrams, makes the new set of representations more functional for the day-to-day work of users.

There are still some limitations to overcome in the newly proposed DEMO representations, duly identified in section III-G, but most are or will be mitigated with functionalities we are developing in our modelling tool.

Regarding our results, incorporating a hands-on exercise to evaluate user performance (e.g., speed, ease of use, intuitiveness) with different versions of diagrams should strengthen them. Depending on the users' availability, a real-life situation simulation would help to evaluate task completion using different Process and Fact Models' representations. This family of studies can be enhanced with constructs regarding effectiveness and quality of work.

From our literature review we identify other variables that can be explored in future work [24], [29], namely: cognitive effectiveness (the speed, ease, and accuracy with which the user perceives the diagram), cognitive fit (the use of different visual dialects when required), semiotic clarity (each graphical symbol may only be associated to one semantic construct), perceptual discriminability (symbols should be smoothly distinguishable), semantic transparency (symbols look should be suggestive of its meaning), graphic economy (keep the number of different symbols cognitively manageable), dual coding (enrich diagrams with textual descriptions), physical quality (persistence, currency, and availability of the diagram), syntactic quality (the correct usage of the modeling language), semantic quality (relationship between the diagram and the modeling domain), perceived semantic quality (relationship between the knowledge of the stakeholders and their interpretation of the diagram), pragmatic quality (interpretation of the diagram by technical actors), social quality (relationship between the interpretation of the process modeled by different actors), deontic quality (contribution of the diagram to fulfill the overall goals of modeling), and comprehensiveness (degree to which the diagram represents the complete process that it is meant to represent).

These variables can be used to evaluate and improve the visual notation of diagrams and other representations of DEMO. The evaluation of user experience should be included in any modeling project to assess how representations are perceived and to investigate refinements.

Future work should address more diverse domains and demographics. The questionnaire could be extended to include open-ended questions that directly ask participants for improvements in each representation and explore the reasons behind their preferences. Usability is related to the evaluation of the effectiveness, efficiency, and satisfaction of users working with a system and, therefore, cannot be fully applied in the evaluation of model representations. However, inspired by Nielsen's Usability Model, we suggest that future work should include more objective metrics, such as success rate, time required to complete a task, and error rate (including critical errors and non-critical errors). Also, an expert-based evaluation of the representations could be added to the study to increase the reliability of the findings.

Owing to a lack of time and resources, and to avoid having a long questionnaire that could lead to user fatigue, our work focused only on the EQ and SPQ constructs under the SEQUAL framework. Future research could conduct a technical evaluation of representations using other metrics established in SEQUAL to identify areas of improvement. This can be complemented by studies that investigate the cognitive effectiveness of representations to understand their impact on user comprehension, task completion, and user learning outcomes. This could be complemented with instruments to assess the cognitive load (e.g., NASA Task Load Index).

ACKNOWLEDGMENT

Special thanks to the participants of these studies who contributed with their fruitful insights and feedback. The authors are grateful to Valentim Caires and Andrea Loreto-Dias for their help in drafting this article and also to the reviewers for their insightful comments, which helped to improve this manuscript. They acknowledge that Grammarly, LanguageTool, and Paperpal Freeflight were used for paraphrasing, grammar, and style checks.

REFERENCES

[1] J. L. G. Dietz and H. B. F. Mulder, *Enterprise Ontology: A Human-Centric Approach to Understanding the Essence of Organisation*. Cham, Switzerland: Springer, 2020, doi: 10.1007/978-3-030-38854-6.

[2] J. Mendling, H. A. Reijers, and W. M. P. van der Aalst, "Seven process modeling guidelines (7PMG)," *Inf. Softw. Technol.*, vol. 52, no. 2, pp. 127–136, Feb. 2010, doi: 10.1016/j.infsof.2009.08.004.

[3] J. Becker, M. Rosemann, and C. von Uthmann, "Guidelines of business process modeling," in *Business Process Management: Models, Techniques, and Empirical Studies*. Berlin, Germany: Springer, 2000, pp. 30–49, doi: 10.1007/3-540-45594-9_3.

[4] D. T. Avila, R. I. dos Santos, J. Mendling, and L. H. Thom, "A systematic literature review of process modeling guidelines and their empirical support," *Bus. Process Manage. J.*, vol. 27, no. 1, pp. 1–23, Jan. 2021, doi: 10.1108/bpmj-10-2019-0407.

[5] H. Leopold, H. van der Aa, F. Pittke, M. Raffel, J. Mendling, and H. A. Reijers, "Integrating textual and model-based process descriptions for comprehensive process search," in *Enterprise, Business-Process and Information Systems Modeling*. Cham, Switzerland: Springer, 2016, pp. 51–65, doi: 10.1007/978-3-319-39429-9_4.

[6] I. M.-M. de Oca, M. Snoeck, H. A. Reijers, and A. Rodríguez-Morffi, "A systematic literature review of studies on business process modeling quality," *Inf. Softw. Technol.*, vol. 58, pp. 187–205, Feb. 2015, doi: 10.1016/j.infsof.2014.07.011.

[7] M. A. T. Mulder, R. Mulder, and F. Bodnar, "Towards a DEMO description in simplified notation script," in *Advances in Enterprise Engineering XVI*, vol. 473. Cham, Switzerland: Springer, 2023, pp. 53–70.

[8] I. Mulder and M. A. T. Mulder, "The effect of nudges, visual literacy and their interaction on enhancing the understanding of enterprise process models," in *Advances in Enterprise Engineering XVI*. Cham, Switzerland: Springer, 2023, pp. 106–120, doi: 10.1007/978-3-031-34175-5_7.

[9] K. Krippendorff, *The Semantic Turn: A New Foundation for Design*, 1st ed., Boca Raton, FL, USA: CRC Press, Dec. 2005, p. 368, doi: 10.1201/9780203299951.

[10] D. Pinto, D. Aveiro, D. Pacheco, B. Gouveia, and D. Gouveia, "Validation of demo's conciseness quality and proposal of improvements to the process model," in *Advances in Enterprise Engineering XIV*. Cham, Switzerland: Springer, 2021, pp. 133–152, doi: 10.1007/978-3-030-74196-9_8.

[11] B. Gouveia, D. Aveiro, D. Pacheco, D. Pinto, and D. Gouveia, "Fact model in DEMO—Urban law case and proposal of representation improvements," in *Advances in Enterprise Engineering XIV*. Cham, Switzerland: Springer, 2021, pp. 173–190, doi: 10.1007/978-3-030-74196-9_10.

[12] J. L. G. Dietz and H. B. F. Mulder, "The MU theory: Understanding models and modelling," in *Enterprise Ontology*. Cham, Switzerland: Springer, Jan. 2020, pp. 71–94, doi: 10.1007/978-3-030-38854-6_6.

[13] Object Management Group. (Jan. 2014). *Business Process Model and Notation (BPMN), Version 2.0.2*. Standard Specification, Object Management Group Standard. [Online]. Available: https://www.omg.org/spec/BPMN/2.0.2

[14] E. F. Codd, "A relational model of data for large shared data banks," *Commun. ACM*, vol. 13, no. 6, pp. 377–387, Jun. 1970, doi: 10.1145/362384.362685.

[15] P. P.-S. Chen, "The entity-relationship model—Toward a unified view of data," *ACM Trans. Database Syst.*, vol. 1, no. 1, pp. 9–36, Mar. 1976, doi: 10.1145/320434.320440.

[16] J. Krogstie, "SEQUAL specialized for business process models," in *Quality in Business Process Modeling*. Cham, Switzerland: Springer, 2016, pp. 103–138, doi: 10.1007/978-3-319-42512-2_3.

[17] N. Genon, D. Amyot, and P. Heymans, "Analysing the cognitive effectiveness of the UCM visual notation," in *System Analysis and Modeling: About Models*, vol. 6598. Berlin, Germany: Springer, 2011, pp. 221–240, doi: 10.1007/978-3-642-21652-7_14.

[18] D. Pacheco, D. Aveiro, B. Gouveia, and D. Pinto, "Evaluation of the perceived quality and functionality of fact model diagrams in DEMO," in *Advances in Enterprise Engineering XV*, vol. 3, Jan. 2022, pp. 114–128, doi: 10.1007/978-3-031-11520-2_8.

[19] D. Pacheco, D. Aveiro, D. Pinto, and B. Gouveia, "Towards the X-theory: An evaluation of the perceived quality and functionality of demo's process model," in *Advances in Enterprise Engineering XV*. Cham, Switzerland: Springer, 2022, pp. 129–148, doi: 10.1007/978-3-031-11520-2_9.

[20] D. Aveiro, V. Freitas, D. Pacheco, and D. Pinto, "Evaluating the perceived quality and functionality of DEMO models' representations in the health domain," in *Proc. 15th Int. Joint Conf. Knowl. Discovery, Knowl. Eng. Knowl. Manage.*, 2023, pp. 331–338, doi: 10.5220/0012260400003598.

[21] *Systems and software engineering—Systems and Software Quality Requirements and Evaluation (SQuaRE)—Product quality Model*, Standard ISO/IEC 25010:2023, International Organization for Standardization, 2023. [Online]. Available: https://www.iso.org/standard/78176.html

[22] J. H. Larkin and H. A. Simon, "Why a diagram is (sometimes) worth ten thousand words," *Cognit. Sci.*, vol. 11, no. 1, pp. 65–100, Jan. 1987, doi: 10.1016/S0364-0213(87)80026-5.

[23] H. Zeevat, "A compositional approach to discourse representation theory," *Linguistics Philosophy*, vol. 12, no. 1, pp. 95–131, Feb. 1989, doi: 10.1007/bf00627399.

[24] D. Moody, "The 'physics' of notations: Toward a scientific basis for constructing visual notations in software engineering," *IEEE Trans. Softw. Eng.*, vol. 35, no. 6, pp. 756–779, Nov. 2009, doi: 10.1109/TSE.2009.67.

[25] W. Sedera, M. Rosemann, and G. Doebeli, "A process modelling success model: Insights from a case study," in *Proc. 11th Eur. Conf. Inf. Syst.*, Jan. 2003, pp. 1–11. [Online]. Available: https://eprints.qut.edu.au/11129/1/11129.pdf

[26] R. E. Mayer, "The past, present, and future of the cognitive theory of multimedia learning," *Educ. Psychol. Rev.*, vol. 36, no. 1, pp. 1–25, Mar. 2024, doi: 10.1007/s10648-023-09842-1.

[27] J. Sweller, "Discussion of the special issue on cognitive load theory," *Brit. J. Educ. Psychol.*, vol. 93, no. S2, pp. 402–410, Aug. 2023, doi: 10.1111/bjep.12606.

[28] F. Paas and J. J. G. van Merriënboer, "Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks," *Current Directions Psychol. Sci.*, vol. 29, no. 4, pp. 394–398, Aug. 2020, doi: 10.1177/0963721420922183.

[29] J. Krogstie, "Quality of business process models," in *Quality in Business Process Modeling*, J. Krogstie, Ed., Cham, Switzerland: Springer, 2016, pp. 53–102, doi: 10.1007/978-3-319-42512-2_2.

[30] J. Krogstie, *Model-Based Development and Evolution of Information Systems: A Quality Approach*. Cham, Switzerland: Springer, 2012, doi: 10.1007/978-1-4471-2936-3.

[31] J. Mendling, H. A. Reijers, and J. Recker, "Activity labeling in process modeling: Empirical insights and recommendations," *Inf. Syst.*, vol. 35, no. 4, pp. 467–482, Jun. 2010, doi: 10.1016/j.is.2009.03.009.

[32] *Systems and Software Engineering—Systems and Software Quality Requirements and Evaluation (SQuaRE)–Measurement Reference Model and Guide*, Standard ISO/IEC 25020:2019, International Organization for Standardization, 2019. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso-iec:25020:ed-2:v1:en

[33] *Systems and Software Engineering—Systems and Software Quality Requirements and Evaluation (SQuaRE)—Evaluation Process*, Standard ISO/IEC 25040:2024, International Organization for Standardization, 2024. [Online]. Available: https://www.iso.org/obp/ui/en/#iso:std:iso-iec:25040:ed-2:v1:en

[34] A. R. Hevner, "A three cycle view of design science research," *Scandin. J. Inf. Syst.*, vol. 19, no. 2, p. 4, Jan. 2007. [Online]. Available: https://aisel.aisnet.org/sjis/vol19/iss2/4

[35] V. Freitas, D. Pinto, V. Caires, L. Tadeu, and D. Aveiro, "The DISME low-code platform—From simple diagram creation to system execution," in *Proc. 22nd CIAO! Doctoral Consortium, Enterprise Eng. Working Conf. Forum*, vol. 3388, Nov. 2022. [Online]. Available: https://ceur-ws.org/Vol-3388/#paper4

[36] V. R. Basili, F. Shull, F. Lanubile, and I. C. Society, "Building knowledge through families of experiments," *Knowl. Creation Diffusion Utilization*, vol. 25, no. 4, pp. 456–473, 1999, doi: 10.1109/32.799939.

[37] D. Aveiro and D. Pinto, "Interview based demo axioms' benefits validation: A regional government case study," in *Knowledge Discovery, Knowledge Engineering and Knowledge Management*, A. Fred, J. L. G. Dietz, D. Aveiro, K. Liu, and J. Filipe, Eds., Cham, Switzerland: Springer, 2015, pp. 421–443, doi: 10.1007/978-3-319-25840-9_26.

[38] D. Aveiro and D. Pinto, "Devising demo guidelines and process patterns and validating comprehensiveness and conciseness," in *Enterprise, Business-Process and Information Systems Modeling*, I. Bider, K. Gaaloul, J. Krogstie, S. Nurcan, H. A. Proper, R. Schmidt, and P. Soffer, Eds., Berlin, Germany: Springer, 2014, pp. 408–423, doi: 10.1007/978-3-662-43745-2_28.

[39] E. Wang and B. S. Caldwell, "An empirical study of usability testing: Heuristic evaluation vs. user testing," in *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, Sep. 2002, vol. 46, no. 8, pp. 774–778, doi: 10.1177/154193120204600802.

[40] D. M. Mertens, *Research and Evaluation in Education and Psychology: Integrating Diversity With Quantitative, Qualitative, and Mixed Methods*, 6th ed., Newbury Park, CA, USA: Sage, 2024. [Online]. Available: https://collegepublishing.sagepub.com/products/research-and-evaluation-in-education-and-psychology-6-277790

[41] D. Pacheco, D. Aveiro, D. Pinto, B. Gouveia, and V. Freitas, "User-perceived quality and functional suitability in DEMO's process and fact models," *IEEE Dataport*, May 2024, doi: 10.21227/zvhm-8t25.

[42] A. P. Field, *Discovering Statistics Using IBM SPSS Statistics*. Los Angeles, CA, USA: Sage, Feb. 2024. [Online]. Available: https://collegepublishing.sagepub.com/products/discovering-statistics-using-ibm-spss-statistics-5-260423

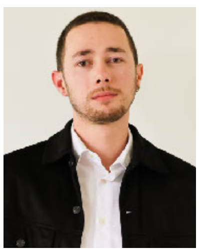
[43] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslèn, *Planning*. Berlin, Germany: Springer, 2024, pp. 115–145, doi: 10.1007/978-3-662-69306-3_9.

DULCE PACHECO was born in Funchal, Madeira, Portugal. She received the B.A. degree in organizational communication from the Polytechnic Institute of Coimbra, Portugal, the M.A. degree in human resources management from the Polytechnic University of Madrid, Spain, and the Ph.D. degree in psychology from the University of Madeira, Portugal, focused on collaborative learning.

She is currently a Research Fellow at ARDITI, Portugal, specializing in process modeling and organizational change. She is focused on the EDIH Smart Islands Hub and on digital transformation initiatives within small and medium enterprises. Besides this, she teaches human resources management and organizational behavior courses at the University of Madeira. After earning her bachelor's degree, she spent over a decade in management and organizational change roles, primarily within educational settings. She further expanded her knowledge base through postdoctoral grants and research fellowships, which deepened her research interests to encompass their current breadth. Therefore, she has a blend of hands-on experience and academic achievement in her career path. Her current research explores relevant areas, such as organizational behavior, enterprise engineering, organizational change, human factors, and collaborative learning.

DAVID AVEIRO was born in Portugal, in 1979. He received the Ph.D. degree in computer science and information systems engineering from the Instituto Superior Técnico, University of Lisbon, in 2010. He is currently an Assistant Professor at the University of Madeira, where he is the Coordinator of the Department of Informatics Engineering and Interactive Media Design. He is also a Researcher at Madeira's Regional Agency for the Development of Research, Technology and Innovation (ARDITI) and at the NOVA LINCS Laboratory-Madeira University Branch. He teaches undergraduate and master's degrees in informatics engineering in the areas of enterprise engineering, information systems, database management systems, and decision-support systems. He has approximately 90 publications, including conference and journal papers, book chapters, and edited books. He was the Program Chair in 30 events as well as a Supervisor of eight Ph.D. theses and 40 M.Sc./B.Sc. projects. He also participates and/or participated as the Principal Investigator in 13 projects competitively funded by European, National, or Regional programs. His research interests include enterprise engineering and digital innovation. He is a member of the management board of The Netherlands-based Enterprise Engineering Institute, which is responsible for the maintenance of the DEMO standard.

DUARTE PINTO was born in Funchal, Madeira, Portugal, in 1987. He received the B.S. and M.S. degrees in software engineering from the University of Madeira, in 2009 and 2013, respectively. He has been certified as a FIWARE Expert, since 2018. Currently, he is a Researcher at the Enterprise Engineering Laboratory (EELab), ARDITI, where he has participated in multiple projects (three national and six international). His research domains include enterprise engineering, open-source smart solutions, information systems, e-government, open data, data spaces, and FIWARE. He is the author/co-author of more than 20 articles in these areas.



VÍTOR FREITAS was born in Madeira, Portugal, in 1997. He received the B.S. and M.S. degrees in software engineering from the University of Madeira, Portugal, in 2020 and 2023, respectively. He has been with the Enterprise Engineering Laboratory (EELab), since 2020, as a Software Engineer. Although, he is still in the early stages of his career and he is the author/co-author of five articles.