Experience Evaluation Report for DSLs for Clinical Protocols: The Case of GDL and PROforma DSLs from a Medical Perspective

Godwin Ukeje[1], Sofia Meacham[1] and Simon DS Fraser[2]

[1]Bournemouth University, Poole, UK
[2]School of Primary Care, Population Sciences and Medical Education, Faculty of Medicine, University of Southampton, Southampton, UK
smeacham@bournemouth.ac.uk

Abstract. This study evaluates the effectiveness of two Domain-Specific Languages (DSLs) — the Guidelines Definition Language (GDL) and PROforma — in developing clinical protocols from the perspective of medical professionals. The objective was to assess the usability, learnability, and overall performance of these tools through hands-on tasks and paper exercises. Participants, including general practitioners, nephrologists, and public health specialists, provided feedback on the GDL 2 Editor and PROforma DSL based on their experience in creating clinical protocols. GDL 2 Editor was found to be more intuitive and user-friendly for simpler tasks, though it faced challenges in handling more complex logic and conditional rules. In contrast, PROforma, despite being designed for advanced decision-making, suffered from usability issues due to its technical complexity and the absence of an Integrated Development Environment (IDE), which limited its practical application. The results highlight the need for further refinement of both tools, with a focus on improving IDE functionality, enhancing user-centered design, and integrating the DSLs more seamlessly into clinical systems to support widespread adoption in healthcare settings. Future work will concentrate on improving usability, adding support tools, and ensuring practical applicability to meet the evolving needs of medical professionals.

Keywords: Domain-Specific Languages (DSLs), Clinical Protocols, Guidelines Definition Language (GDL), PROforma, Usability, Integrated Development Environment (IDE), Clinical Decision Support Systems (CDSS), Healthcare IT Integration, User-Centered Design, Medical Informatics.

1 Introduction

Clinical Decision Support Systems (CDSS) have become integral tools in modern healthcare, providing real-time guidance to clinicians by facilitating the development and application of clinical protocols. These protocols support standardization in medical care, reducing variability in treatment and improving patient outcomes [1]. CDSS adoption has been widely acknowledged as a critical step toward improving healthcare delivery, with emphasis on addressing usability and integration challenges [2].

In this context, at the heart of Clinical Decision Support Systems (CDSSs) lie the clinical protocols. Clinical protocols are typically presented in the form of structured or unstructured documents, such as guidelines, care pathways, and recommendations such as NICE guidelines in the UK [3]. To transform these textual protocols into actionable tools, the information must be extracted, formalized, and computerised. This digital representation becomes the core of CDSSs, which provide automated reasoning and decision-making support to healthcare professionals. Transforming theoretical knowledge of clinical concepts into actionable steps for diagnosing and managing diseases is a complex task [4]. Failure to adhere to clinical guidelines and protocols can pose significant risks to patients, such as delays in identifying conditions like CKD [5] or inadequate management of critical factors like hypertension [6]. Moreover, healthcare professionals must continuously adapt to evolving clinical guidelines and new research findings, both during their initial training and throughout their ongoing professional development [7].

Domain-Specific Languages (DSLs) play a pivotal role in this process by offering a specialized framework to encode clinical protocols. DSLs are designed to simplify and streamline the modeling of domain knowledge [8]. This could be complex medical workflows, ensuring precision, consistency, and usability. By translating the intricate details of clinical guidelines into executable programs, DSLs enable the structured and accurate implementation of these protocols. This approach not only enhances the maintainability of clinical decision support systems but also ensures the alignment of CDSS functionalities with evidence-based guidelines. DSLs thereby act as a bridge between documented clinical knowledge and its practical, computerised application, forming the foundation of reliable and effective healthcare support systems.

This paper focuses on the evaluation of two prominent DSLs: the Guidelines Definition Language (GDL) [9] and PROforma [10]. Both languages are tailored for the healthcare domain and allow medical professionals to develop and apply clinical protocols within CDSS. The GDL has been in use for over a decade, with its development beginning in the early 2010s. GDL, part of the openEHR framework, leverages archetypes—reusable models of clinical concepts—to ensure standardization and accuracy in protocol development [11]. PROforma has been in use for over 25 years, having been developed in the mid-1990s by a research team led by John Fox at Cancer Research UK. PROforma is a formal language designed to represent clinical decisions, plans, actions, and enquiries, emphasizing “correct-by-construction” solutions that minimize errors in medical protocols [10, 12, 13]. Please note that these are not the only DSLs available for clinical protocols; rather, they were selected based on their significant impact and prominence in the field.

While DSLs hold great potential in the medical field, their usability and practical application present significant challenges. Medical professionals, who are the end-users of these systems and often very busy with clinical tasks, often lack the technical expertise required to fully utilize such tools, especially when dealing with complex syntax or development environments. In addition, the integration of DSLs into existing healthcare systems can be cumbersome, particularly without user-friendly Integrated Development Environments (IDEs) [14, 15]. Therefore, it is crucial to evaluate the usability and learnability of these DSLs to ensure their successful adoption in clinical practice.

The development of DSLs such as GDL and PROforma represents an innovative step in the effort to standardise and streamline clinical decision-making processes. However, their practical implementation highlights a critical problem: the gap between the technical complexity of these languages and the usability requirements of healthcare professionals. As clinicians are not typically trained in programming or advanced technical operations, there is a pressing need for tools that are not only accurate and reliable but also intuitive and easy to learn. This usability gap presents a significant barrier to the broader adoption of DSLs in routine clinical practice, potentially limiting their impact on healthcare outcomes [16].

Furthermore, the integration of DSLs into existing healthcare systems remains a significant challenge. Many healthcare facilities operate with heterogeneous IT infrastructures that lack compatibility with advanced DSL-driven CDSS. Without seamless integration, the potential benefits of DSLs—such as reducing variability in treatment and improving adherence to evidence-based guidelines—cannot be fully realised.

This research aims to generate experiential evidence by evaluating the two most prominent DSLs in the field, incorporating feedback from medical professionals. The focus is on assessing the DSLs ability to address the practical needs of clinicians while ensuring accurate and reliable clinical decision support [2]. The outcome of this evaluation will form the basis for future DSL developments in the field and/or improvements in the current prominent DSLs. The most important achievement will be the direct evaluation of DSLs from medical professionals which can be a rare case for DSLs.

The remainder of this paper is structured as follows: Section 2 presents a review of related work in the field of DSL evaluation in healthcare. Section 3 outlines the experiment design and methodology. Section 4 discusses the evaluation results, focusing on the usability and practical challenges encountered by participants. Finally, Section 5 concludes the paper and outlines future work to enhance the evaluated DSLs for clinical protocols and pave the way for improved and new developments to address the medical professionals requirements.

2 Related Work

In this section, we will first explore the existing research on the evaluation of DSLs, both in general and specifically within the healthcare domain. This will provide a foundation for understanding the methodologies and criteria used to assess the effectiveness of DSLs. Following this, we will delve into the unique challenges associated with DSLs for clinical protocols in healthcare, outlining the specific issues they aim to address and the methods employed to evaluate their performance and utility in real-world clinical settings.

2.1 Evaluating DSLs: General Approaches

Evaluating Domain-Specific Languages (DSLs) involves a variety of methods aimed at assessing their effectiveness, usability, and correctness across different domains.

These approaches typically focus on key aspects such as correctness assessment, functional suitability, usability, and performance evaluation, each contributing to a holistic understanding of DSL capabilities.

For example, correctness is a critical factor in specialized fields like threat modeling. Katsikeas et al. in [17] demonstrated this through their evaluation of the coreLang DSL, which outperformed human experts in attack simulations, showcasing its reliability for cybersecurity applications.

Similarly, functional suitability has been explored in domains such as cognitive training, where Leon et al. in [18] applied an ISO-based quality model to evaluate DSLs, emphasizing their effectiveness and efficiency in generating interactive training applications.

Usability evaluation is another vital consideration, as DSLs are intended to simplify development for users with varying levels of technical expertise. However, as Rodrigues et al. in [19] noted, many studies lack objective usability assessments, underscoring the importance of integrating Human-Computer Interaction (HCI) techniques into DSL evaluation frameworks. Evaluating the usability and effectiveness of DSLs is crucial for their successful adoption, particularly in non-technical domains like healthcare. A growing body of research focuses on usability assessment frameworks and techniques for DSLs, which are essential to determine their applicability and user-friendliness in CDSS [20]. One prominent method for usability evaluation is the Think Aloud protocol, which allows real-time insights into user interactions by encouraging participants to verbalize their thought processes as they use a system. This method has proven effective in identifying usability issues and understanding user behaviour [19, 21]. Poltronieri et al. [8] proposed a usability evaluation framework called USA-DSL, specifically designed for DSLs, which emphasizes the need for integrating usability considerations from the early stages of DSL development.

Performance evaluation further highlights DSL advantages, with studies like [22] showing that DSLs such as Athos can significantly enhance task efficiency compared to general-purpose languages, especially for less experienced developers. Challenger et al. in [23] provided a comprehensive framework for evaluating DSLs, incorporating dimensions such as usability, functional quality, and execution correctness. While these evaluation strategies offer valuable insights, challenges remain in standardizing methodologies across diverse applications, which limits broader adoption and understanding of DSLs' benefits.

2.2 Challenges and Evaluation of DSLs for Clinical Protocols in Healthcare

In healthcare, ensuring the correct implementation of clinical protocols is essential to standardize care, minimize variability in treatment, and adhere to evidence-based guidelines. However, these protocols often involve intricate workflows, overlapping decision paths, and the simultaneous application of multiple guidelines, making their integration into clinical practice a challenging task [24]. This complexity is further compounded by potential conflicts between guidelines, which can lead to inconsistencies in decision-making. Addressing these challenges requires robust frameworks capable of managing the intricacies of clinical protocols while optimizing patient outcomes. For instance, Wilk et al. proposed a mitigation framework designed to handle the concurrent application of multiple clinical practice guidelines, highlighting the critical role of decision support systems in reducing conflicts and ensuring cohesive care delivery [25]. Additionally, Morris et al. work on treatment burden and multimorbidity emphasizes the challenges of managing complex care processes and provides insights into improving the alignment of clinical protocols with patient needs [26].

DSLs like GDL and PROforma have been applied to model clinical guidelines, allowing medical professionals to automate routine clinical decisions while ensuring consistency and reducing the likelihood of human error [11]. However, their success largely depends on how well these languages align with the needs and workflows of healthcare practitioners. For example, GDL, built within the openEHR framework, ensures standardization through archetypes, which are reusable models of clinical concepts [11]. PROforma, a more formalized language, provides structured methods for decision-making but faces challenges in practical application without sufficient technical support [15].

Despite their potential, usability remains a significant barrier to the widespread adoption of DSLs in healthcare. Several studies have emphasized the importance of user-centred design in the development of DSLs to ensure that they meet the practical needs of end-users, particularly healthcare professionals who may not have technical expertise [16] [15]. A study [21] highlighted the need for iterative usability testing during DSL development, recommending the involvement of end-users in the design process to refine language features and ensure applicability. Similarly, [15] proposed that DSL usability should be evaluated not just on technical functionality but also on how seamlessly the language integrates into clinical workflows. This includes providing robust Integrated Development Environments (IDEs) that allow medical professionals to easily write, test, and validate clinical protocols without extensive programming knowledge.

Overall, while DSLs such as GDL and PROforma present promising solutions for automating clinical protocols, their usability, technical complexity, and lack of seamless integration with existing healthcare systems pose significant challenges. These limitations hinder their effective application and broader adoption in healthcare settings, underscoring the need for further refinement and user-centred improvements.

Usability and Learnability Issues. One of the most significant limitations of existing DSLs is their usability. Medical professionals, the primary users of these tools, often lack technical expertise in programming languages. PROforma, for instance, while offering robust capabilities for modelling clinical decision pathways, has been criticised for its rigid syntax and steep learning curve [12]. Similarly, the GDL 2 Editor, though more user-friendly, struggles with managing complex protocols due to insufficient support for advanced rule constructions and inadequate guidance on handling conditional logic [8]. These challenges create a usability gap that limits the ability of clinicians to adopt and effectively use these tools without substantial training or technical support.

Integration with Existing Healthcare Systems. A critical issue with current DSLs is their lack of seamless integration with existing electronic health record (EHR) systems. Many DSLs require manual input of patient data or lack mechanisms for real-time data retrieval, which can disrupt clinical workflows [13]. For instance, DSLs often operate as standalone systems rather than being embedded into the IT infrastructure of healthcare facilities, leading to inefficiencies and reduced adoption. Studies have emphasised the importance of developing CDSS tools that integrate smoothly with heterogeneous healthcare IT environments to ensure usability in real-world clinical settings [25].

Scalability and Maintenance Challenges. Another limitation lies in the scalability and maintainability of DSL-based clinical protocols. As medical guidelines evolve, updating existing protocols becomes cumbersome, particularly for languages like PROforma, which lack intuitive interfaces for modifying existing rules [12]. Moreover, the absence of real-time validation features and debugging tools can lead to errors in protocol implementation, which may go undetected until adverse events occur in clinical practice [27]. These shortcomings point to the need for improved tools that allow dynamic updates while maintaining the reliability of clinical decision-making processes.

Lack of User-Centred Design. Despite the growing emphasis on user-centred design in software development, DSLs in healthcare have largely failed to adequately incorporate the perspectives of end-users during their design and development stages. Existing DSLs have often prioritised technical capabilities over ease of use, resulting in tools that are difficult for clinicians to integrate into their workflows [16]. Research indicates that involving medical professionals in the iterative design process is critical for creating tools that align with their practical needs and constraints [2].

Limited Evidence of Real-World Effectiveness. There is a dearth of empirical evidence demonstrating the real-world effectiveness of DSLs in improving clinical outcomes. While theoretical benefits such as increased standardisation and reduced variability in care have been highlighted, few studies have rigorously evaluated the impact of these tools in operational healthcare settings [25]. The absence of such evidence makes it difficult to justify investment in DSLs, particularly in resource-constrained healthcare systems.

This study aims to build on previous research by evaluating these DSLs through a usability-focused approach, emphasizing the practical challenges faced by healthcare professionals in real-world applications. Through a combination of usability assessments and participant feedback, this study contributes to the growing body of literature on how DSLs can be optimized for clinical use, ensuring they are accessible, efficient, and effectively integrated into healthcare practice.

3 Experiment Design

The experiment was designed to evaluate the effectiveness of DSLs for developing clinical protocols, focusing specifically on the GDL and PROforma DSLs. The primary aim was to assess these DSLs in terms of usability, learnability, and applicability from the perspective of medical professionals. A diverse group of 21 participants was selected from the authors’ professional networks and were invited by email, including general practitioner (GPs), nephrologists, public health specialists, and an emergency physician to provide a wide range of insights and experiences.

The recruitment process for participants involved purposive sampling [28], selecting individuals based on their clinical expertise and potential for providing diverse perspectives on the tools. While the authors' professional networks served as the primary recruitment pool, efforts were made to ensure that the selected group represented varied medical disciplines and levels of familiarity with digital tools. The structured email invitations outlined the study objectives, the expected time commitment, and the tasks participants would perform, ensuring informed consent and voluntary participation.

Participants were first introduced to the tools through an 18-minute instructional video and written documentation. This preparatory phase was essential in establishing a baseline understanding of the DSLs' interfaces and functionalities. The instructional materials provided a detailed overview of how to use the GDL 2 Editor and the PROforma DSL, ensuring that participants were well-prepared for the hands-on tasks that followed. Moreover, the instructional materials provided were standardised and validated by pilot testing [29] with a small group of healthcare professionals prior to the main study. The instructional video and documents were designed to ensure uniformity in baseline knowledge across participants, minimising bias caused by differences in prior familiarity with DSLs. These steps aimed to create a controlled and repeatable experimental framework, enabling the authors to draw reliable conclusions about the usability and learnability of the GDL and PROforma DSLs.

After the introductory phase, participants engaged in a task using the GDL 2 Editor, where they were required to create a protocol for Chronic Kidney Disease (CKD) diagnosis based on provided documentation. This involved instantiating archetypes, defining input and output parameters, and constructing rule lists to encapsulate the clinical logic for managing CKD [27]. Archetypes, reusable models of clinical concepts, are central to GDL, ensuring the standardization and accuracy of medical data representation [13]. The task allowed participants to experience the practical challenges of protocol development using GDL and to provide feedback on the tool's usability.

For the PROforma task, participants did not engage in a practical application due to the technical expertise required for using the language. Instead, the task was conducted as a paper exercise, where participants were asked to write code for a clinical protocol in PROforma without testing it. This highlighted the complexities and challenges associated with using PROforma in a real-world setting without adequate technical support or a dedicated IDE [13].

To evaluate the usability of these DSLs, a modified Think Aloud method was employed. Participants were asked to verbalize their thoughts while interacting with the DSLs, offering real-time insights into their cognitive processes and the challenges they encountered [30]. This method was chosen for its ability to uncover usability issues that might not be captured through standard questionnaires alone.

Data collection involved both quantitative and qualitative approaches. Participants completed a detailed questionnaire after the task, designed to gather data on various aspects of the DSLs, including intuitiveness, functionality, documentation support, and overall user satisfaction. Quantitative data were collected using a rating scale from 1 (very difficult) to 5 (very easy), while qualitative data were gathered through open-ended written questions that allowed participants to provide detailed feedback and suggestions for improvement [8] [19] . This combination of qualitative and quantitative data helped paint a picture of the user experience.

The methodology also included observation and documentation of participant interactions with the DSLs. This allowed the research team to capture insights into how participants navigated the tools, how they approached problem-solving, and the specific challenges they faced during the tasks. This approach to data collection enabled a robust evaluation of the DSLs, providing actionable feedback for future improvements.

Overall, the experiment was methodically designed to assess the GDL 2 Editor and PROforma DSL in real-world scenarios, with a strong focus on usability and applicability. By involving medical professionals and collecting both quantitative and qualitative data, the study aimed to inform future refinements of these tools, and the requirements for new DSL developments.

4 Evaluation Results

4.1 Usability and Learnability of GDL 2 Editor

The GDL 2 Editor received relatively positive feedback, particularly in terms of usability and ease of use for general protocol development. Many participants appreciated the archetype system, which allowed for standardization and reuse of clinical concepts. However, despite this, there were clear challenges identified. A significant number of participants (38%) found instantiating archetypes and defining input/output parameters difficult. Several comments pointed out that the GDL 2 Editor lacked detailed guidance on how to structure more complex rule lists, which made the tasks challenging for less technically inclined users.

While the editor provided a framework for rule construction, only 14% of participants reported feeling “very confident” when constructing rule lists, with most feeling either neutral (29%) or slightly confident (33%). Participants expressed concerns about the limited support for conditional rules and the lack of flexibility when modifying archetypes once they had been instantiated. Several respondents also highlighted that navigating between multiple archetypes was cumbersome, and the editor became slower when handling multiple instances. This suggests that while GDL 2 is effective for basic tasks, it struggles with more complex clinical protocols.

Table 1. The average ratings of GDL 2 Editor features

Feature	Average Results (1-5)
Intuitiveness	3.8
Ease of Use	3.6
Supportiveness of Documentation	3.4
Learnability	4.0
Handling of Complex Tasks	2.9
User Interface	3.7
Performance in Simple Tasks	4.2

Flexibility	3.1
Error Management	2.8

Additionally, there was a call for better integration of user feedback within the GDL 2 Editor itself. Many participants recommended adding real-time validation features to help users identify mistakes and ensure that the protocols were correctly constructed before being finalized. The absence of this feature led to uncertainty about the accuracy of the protocols created. In Fig. 1, the overall results for GDL 2 are presented.

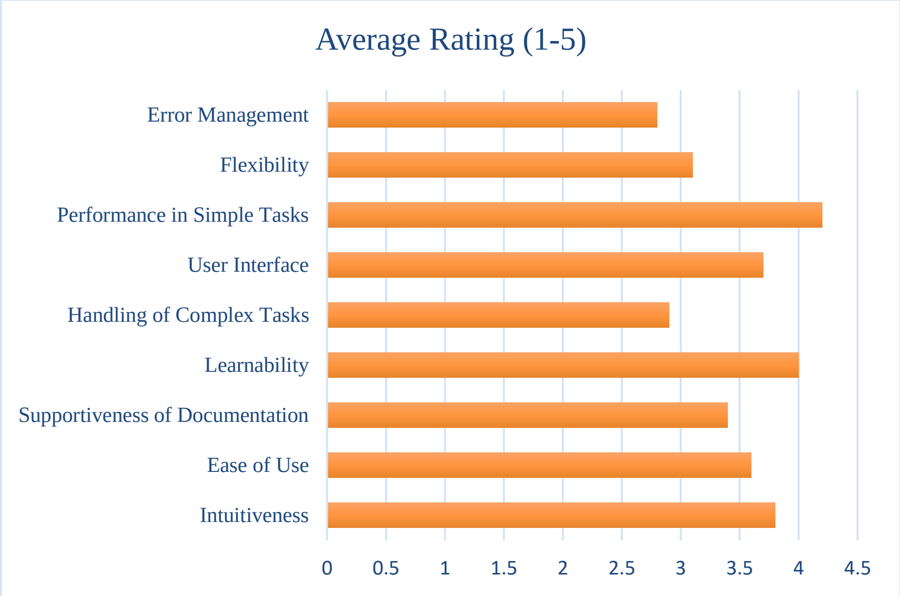

Fig. 1: Summary of GDL 2 Feature Ratings

4.2 PROforma DSL: Paper Exercise Challenges

The evaluation of PROforma posed more significant challenges, largely due to the technical complexity of the DSL and the lack of a dedicated IDE. Unlike GDL, participants did not engage in a practical application of PROforma but were instead asked to complete a paper-based exercise. This highlighted several key limitations.

Most notably, many participants struggled with the rigid syntax of PROforma, with 22% of participants rating the clarity of the syntax as "Not clear at all," and another 22% finding it only "Slightly clear." The lack of real-world examples in the documentation further complicated the task, as many participants found it difficult to apply the theoretical concepts to the clinical decision-making process. The absence of an IDE and debugging tools also contributed to frustration, as participants were unable to test or validate their code, making the task feel more abstract and disconnected from practical clinical workflows.

Participants repeatedly emphasized the need for a visual editor or a drag-and-drop interface that would simplify the process of creating protocols. Additionally, there was strong feedback requesting a built-in testing environment where participants could simulate the clinical logic and verify that their protocols functioned as expected. Without these tools, the participants found it difficult to determine whether their decision logic was correct, which created a significant barrier to usability.

4.3 Comparative Analysis: GDL vs. PROforma

When comparing GDL and PROforma, GDL was generally viewed as more user-friendly, particularly for those without a technical background. GDL's archetype-based system provided a level of structure that made it easier for first-time users to construct basic protocols. 43% of participants found GDL easier to use for clinical protocol development, while 21% selected PROforma, and 14% found both tools equally challenging. GDL was praised for its simplicity and intuitiveness, especially for routine tasks, whereas PROforma was recognized for offering more flexibility and control over complex decision-making processes, though at the cost of usability.

In terms of real-world applicability, 33% of participants felt that GDL was more suitable for clinical practice due to its ease of use and structured approach to rule construction. However, another 33% selected PROforma, citing its adaptability and control over finer details of clinical protocols as advantages, particularly for more complex cases. Despite this, the majority agreed that PROforma would require significant improvements in usability, particularly in the form of an IDE with integrated debugging and testing capabilities before it could be widely adopted in clinical settings.

Table 2. Numerical comparative analysis (mean) of GDL 2 Editor and PROforma DSL

Criteria	**GDL 2 Editor (1-5)**	**PROforma DSL (1-5)**
Practical Application	4.0	2.0
Task Method	4.0	2.0
Intuitiveness and Usability	3.8	2.5
Supportiveness of Materials	3.4	2.3
Learnability	4.0	2.0
Overall Performance	3.9	2.1
Challenges	3.2	3.0
Integration	2.8	1.9
Technical Support	3.6	1.8

The table above provides a numerical comparison across key criteria, with GDL 2 Editor generally receiving higher ratings in most categories, particularly in usability and practical application, while PROforma faced challenges due to its lack of support tools and complexity.

Fig. 2 and Fig. 3 below clearly show that participants rated the GDL 2 Editor significantly higher than PROforma DSL across all categories, including usability, learnability, support, and overall performance.

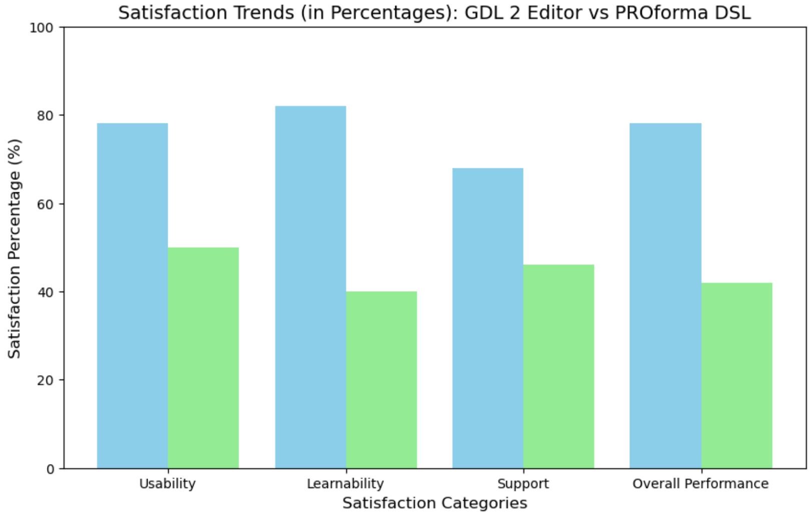

Fig. 2: Satisfaction Trends in Percentages: GDL 2 Editor vs PROforma DSL

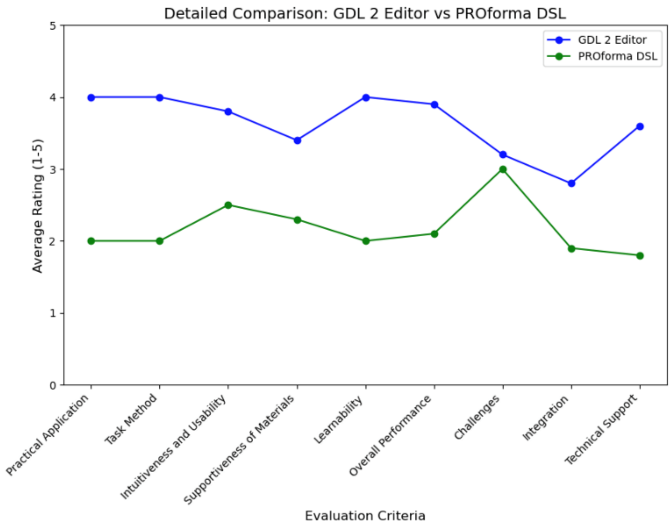

Fig. 3: Detailed Comparison: GDL 2 Editor vs PROforma DSL

The table below highlights the strengths and areas for improvement in both the GDL 2 Editor and PROforma DSL, guiding future developments to enhance their practical application in clinical settings.

Table 3. Summaries of strengths and limitations for the GDL 2 Editor and PROforma DSL

Criteria	**GDL 2 Editor**	**PROforma DSL**
Practical Application	Widely used in the experiment with practical application in building simple protocols.	Participants did not engage in practical application due to the complexity of the tool.
Task Method	Hands-on task creating a clinical protocol using the GDL 2 Editor.	Paper exercise writing code for PROforma; participants were unable to test or validate their work.
Intuitiveness and Usability	Generally found to be intuitive for basic tasks but struggles with complex rule construction.	Not intuitive, especially without a dedicated IDE. Participants found the syntax challenging.
Supportiveness of Materials	Instructional materials were found helpful, but some participants needed more advanced support.	Materials were not sufficient; participants requested more real-world examples and better documentation.
Learnability	Easier to learn for first-time users; archetype-based system helped structure the process.	Difficult to learn without extensive technical knowledge. Steep learning curve due to lack of support tools.
Overall Performance	Rated moderately high for overall performance in simple tasks but lacked flexibility for complex cases.	Overall performance was poor, mainly due to the absence of validation and testing tools.
Challenges	Difficulty in handling multiple archetypes; limited guidance for conditional rules.	Rigid syntax, no error-checking or testing environment, and lack of flexibility in decision paths.
Integration	No seamless integration with other clinical systems; manual input was required.	No integration features provided; very limited application in clinical workflows.
Technical Support	No dedicated IDE but more user-friendly interface compared to PROforma.	Lacked any technical support or IDE, making it difficult to use effectively.

4.4 Addressing Challenges and Potential Solutions

The key challenges identified in the evaluation of both DSLs centred around usability, support features, and the complexity of the tools. Participants repeatedly pointed out that both GDL and PROforma lacked adequate tutorials, detailed documentation, and practical examples. In particular, the need for interactive tutorials or video demonstrations was frequently mentioned. Many participants felt that a hands-on, step-by-step guide would have greatly improved their ability to complete the tasks and understand the tools. This aligns with broader research indicating that user support and feedback mechanisms are critical for DSL adoption, especially for non-technical users [30] [31].

For GDL, improving the interface and navigation between archetypes was a priority for many participants. Several users suggested incorporating features like error-checking, auto-complete for input/output parameters, and enhanced rule construction templates. These suggestions aim to reduce cognitive load and make the process of developing clinical protocols more accessible to users unfamiliar with DSLs, supporting findings from usability studies that stress the importance of user-centric design and minimizing complexity [31].

PROforma, on the other hand, faced more fundamental usability challenges. Many participants requested the development of a dedicated IDE that would provide real-time feedback, error-checking, and a testing environment. Such tools would make the DSL more accessible and allow users to confidently write and validate complex decision-making protocols. The inclusion of a visual workflow system, which would allow users to construct protocols graphically rather than through code, was also frequently recommended [31] [32].

Overall, while both GDL and PROforma offer valuable tools for developing clinical protocols, their usability for non-technical users remains a significant challenge. GDL was generally favoured for its simplicity and structured approach, but it still requires improvements in terms of user support and flexibility. PROforma, despite its power, adaptability and formality offered, was hindered by its complexity and lack of user-friendly tools. To make these DSLs more accessible to healthcare professionals, future developments should focus on enhancing user interfaces, providing tutorials, and integrating dedicated development environments that support real-time feedback and testing [30] [33].

4.5 Validation of Findings

To validate the findings of this study, the results were compared with existing literature and usability benchmarks for domain-specific languages (DSLs) in healthcare. The usability ratings for GDL and PROforma were aligned with previous evaluations, such as the findings by Poltronieri et al. [8], which highlighted the importance of user-centred design in DSLs. Specifically, the learnability score of 4.0 for GDL 2 Editor corroborates prior research that positions archetype-based systems as more accessible for clinicians due to their structured approach [12].

Additionally, the challenges identified in the usability and integration of PROforma are consistent with earlier reports, including Fox et al. [12], which emphasised the need for dedicated Integrated Development Environments (IDEs) to support complex rule constructions. This study's findings extend this understanding by providing empirical evidence of these challenges during protocol development tasks.

Benchmarking against Nielsen’s usability heuristics [30] further supports the conclusions. For instance, GDL's moderate performance in error management (2.8) aligns with heuristic principles that highlight error prevention and recovery as critical areas for improvement. Similarly, PROforma’s lack of feedback mechanisms is a clear deviation from usability best practices, further validating the need for an IDE and real-time debugging tools.

The feedback from participants also validated the experimental design, with 86% agreeing that the tasks accurately reflected real-world clinical challenges. This alignment between study design and participant perception ensures that the results are both robust and representative of actual clinical workflows.

5 Biases and Reflections

In evaluating the GDL 2 Editor and PROforma DSL, several potential biases may have influenced the results. First, the study relied on participants with varying levels of technical expertise and experience in clinical protocol development, which may have impacted their perceptions of the tools. Additionally, our recruitment strategy was limited to professional networks, which restricted the number of participants. In the future, a broader recruitment strategy could significantly increase participant numbers. Participants with more technical backgrounds might have found PROforma more manageable, while those with limited technical experience may have rated GDL 2 Editor more favourably due to its relative simplicity. This variance in participant backgrounds introduces bias related to user familiarity with DSLs and clinical decision support systems.

Additionally, the decision to use a paper-based exercise for PROforma due to the lack of a dedicated IDE may have skewed the results against it. Participants were unable to fully experience PROforma's potential through practical application, which may have contributed to the lower satisfaction ratings. Future studies should seek to minimize this bias by providing a more complete toolset, such as a functional IDE, to give participants a fairer experience of both DSLs.

Another source of bias stems from the subjective nature of usability and satisfaction ratings. While quantitative ratings provide valuable insights, they are influenced by individual preferences, experiences, and frustrations encountered during the evaluation process. For instance, participants may have rated a tool more harshly if they encountered technical difficulties or if the learning curve was steeper than expected.

Reflection on the methodology also reveals the importance of real-world testing environments. The lack of integration between the DSLs and clinical systems meant that participants could not fully appreciate the practical applicability of the tools. This highlights the need for future evaluations to include real-world testing scenarios, where users can experience the tools in the context of their actual work environments.

In reflecting on the study, it is evident that enhancing the IDEs, refining user-centred design, and improving integration into healthcare systems will be critical in addressing these biases and ensuring that future evaluations capture a more balanced and accurate assessment of the DSLs. Last but not least, the non-technical users should feel comfortable to change and maintain the DSL, therefore empowering them to use complex technological developments.

6 Conclusions and Future Work

The evaluation of the GDL 2 Editor and PROforma DSL revealed distinct advantages and limitations in their application to clinical protocol development. GDL 2 Editor was generally more user-friendly, with participants appreciating its structured approach and intuitive design for basic tasks. However, it struggled with handling complex protocols and lacked flexibility when dealing with conditional rules. In contrast, PROforma, although powerful in theory, was hindered by its technical complexity, lack of a dedicated IDE, and steep learning curve, which prevented practical use in this evaluation.

Future work should prioritize the development of robust IDEs for both tools, particularly for PROforma. A dedicated IDE would significantly enhance its usability by offering real-time feedback, error-checking, and debugging tools, making the language more accessible to non-technical users such as medical professionals. For GDL, the focus should be on improving the handling of complex rule sets and adding support for conditional logic to enable users to build more sophisticated protocols without extensive technical expertise.

Another critical area for future research is the adoption of a user-centred design approach. Both DSLs should be refined with an emphasis on simplifying workflows, providing clear guidance, and incorporating user feedback into the development process. By doing so, the tools can better align with the practical needs and capabilities of healthcare professionals, improving the overall user experience and reducing the cognitive load associated with using the systems.

Integration with clinical systems such as those providing electronic health records (EHRs) should also be a key focus of future development. Seamless integration would enhance the practical utility of these DSLs by allowing protocols developed with GDL and PROforma to be implemented and tested in real-world clinical settings. This would not only improve workflow efficiency but also ensure that clinical decision support systems (CDSS) are directly applicable in daily medical practice.

In conclusion, while GDL 2 Editor and PROforma DSL show promise for the development of clinical protocols, their usability, technical support, and integration into clinical systems need substantial improvement. Future research should focus on enhancing IDE functionality, adopting user-centred design principles, and ensuring seamless integration with healthcare IT systems to enable these tools to reach their full potential in clinical practice. Additionally, this experience report outlines and details the requirements for developing new DSLs that address the critical aspects of usability and maintainability, specifically for non-technical users.

References

1. Berner, E.S.: Clinical Decision Support Systems: State of the Art. AHRQ publication. 90069, 1--26 (2009).
2. Greenes, R.: Clinical Decision Support: The Road to Broad Adoption. Academic Press (2014).
3. NICE guidelines | NICE guidance | Our programmes | What we do | About, https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-guidelines, last accessed 2024/11/26.

4. Fraser, S.D., Blakeman, T.: Chronic kidney disease: identification and management in primary care. Pragmatic Obs. Res. 7, 21–32 (2016). https://doi.org/doi: 10.2147/POR.S97310.
5. Fraser, S.D., Parkes, J., Culliford, D., Santer, M., Roderick, P.J.: Timeliness in chronic kidney disease and albuminuria identification: a retrospective cohort study. BMC Fam. Pract. 16, 1–10 (2015). https://doi.org/doi: 10.1186/S12875-015-0235-8/TABLES/3.
6. Fraser, S.D., Roderick, P.J., Mcintyre, N.J., Harris, S., Mcintyre, C.W., Fluck, R.J., Taal, M.W.: Suboptimal blood pressure control in chronic kidney disease stage 3: baseline data from a cohort study in primary care. BMC Fam. Pract. 14, 1–11 (2013). https://doi.org/doi: 10.1186/1471-2296-14-88.
7. Pereira, V.C., Silva, S.N., Carvalho, V.K., Zanghelini, F., Barreto, J.O.: Strategies for the implementation of clinical practice guidelines in public health: an overview of systematic reviews. Health Res. Policy Syst. 20, 13 (2022). https://doi.org/doi: 10.1186/S12961-022-00815-4/TABLES/4.
8. Poltronieri, I., Zorzo, A.F., Bernardino, M., Campos, M. de B.: Usa-dsl: usability evaluation framework for domain-specific languages. In: Proceedings of the 33rd Annual ACM Symposium on Applied Computing. pp. 2013--2021 (2018). https://doi.org/DOI:10.1145/3167132.3167348.
9. Guideline Definition Language (GDL) – Assisting health care personnel and patients, https://gdl-lang.org/, last accessed 2024/11/25.
10. Patkar, V., Fox, J.: Clinical Guidelines and Care Pathways: A Case Study Applying PROforma Decision Support Technology to the Breast Cancer Care Pathway. In: Computer-based Medical Guidelines and Protocols: A Primer and Current Trends. pp. 233–242. IOS Press (2008). https://doi.org/10.3233/978-1-58603-873-1-233.
11. Beale, T., Heard, S.: Archetype Definitions and Principles, https://specifications.openehr.org/releases/1.0.2/architecture/am/archetype_principles.pdf, (2007).
12. Fox, J., Johns, N., Lyons, C., Rahmanzadeh, A., Thomson, R., Wilson, P.: PROforma: a general technology for clinical decision support systems. Comput. Methods Programs Biomed. 54, 59–67 (1997). https://doi.org/10.1016/S0169-2607(97)00034-5.
13. Fox, J., Johns, N., Rahmanzadeh, A.: Disseminating medical knowledge: the PRO*forma* approach. Artif. Intell. Med. 14, 157–182 (1998). https://doi.org/10.1016/S0933-3657(98)00021-9.
14. Borum, H.S., Niss, H., Sestoft, P.: On Designing Applied DSLs for Non-Programming Experts in Evolving Domains. In: 2021 ACM/IEEE 24th International Conference on Model Driven Engineering Languages and Systems (MODELS). pp. 227–238 (2021). https://doi.org/10.1109/MODELS50736.2021.00031.
15. Peleg, M., Tu, S., Bury, J., Ciccarese, P., Fox, J., Greenes, R., Hall, R., Johnson, P., Jones, N., Kumar, A., others: Comparing Computer-interpretable Guideline Models: A Case-study Approach. J. Am. Med. Inform. Assoc. 10, 52--68 (2003). https://doi.org/DOI:10.1197/jamia.m1135.
16. Hersh, W.R., Gorman, P., Biagioli, F., Mohan, V., Gold, J., Mejicano, G.: Beyond information retrieval and electronic health record use: competencies in clinical informatics for medical education. Adv. Med. Educ. Pract. 205--212 (2014). https://doi.org/DOI:10.2147/amep.s63903.
17. Katsikeas, S., Rencelj Ling, E., Johnson, P., Ekstedt, M.: Evaluating the Correctness of Domain-Specific Threat Modeling Languages. Pontus Ekstedt Mathias Eval. Correctness Domain-Specif. Threat Model. Lang. (2023).
18. León, D.S., Jara, K.M., Prado, D., Mora, B.T., Solís, W.V., Orellana, P.C.: Evaluating functional suitability of domain-specific languages oriented to create interactions for cognitive training. In: 2021 Eighth International Conference on eDemocracy & eGovernment (ICEDEG). pp. 124–130. IEEE (2021).

19. Poltronieri Rodrigues, I., de Borba Campos, M., Zorzo, A.F.: Usability Evaluation of Domain-Specific Languages: A Systematic Literature Review. In: Kurosu, M. (ed.) Human-Computer Interaction. User Interface Design, Development and Multimodality. pp. 522–534. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-58071-5_39.
20. Starr, T.: DSL Advances. Prentice Hall Professional (2003).
21. Bariic, A., Amaral, V., Goulao, M.: Usability Evaluation of Domain-Specific Languages. In: 2012 Eighth International Conference on the Quality of Information and Communications Technology. pp. 342--347. IEEE (2012).
22. Hoffmann, B., Urquhart, N., Chalmers, K., Guckert, M.: An empirical evaluation of a novel domain-specific language–modelling vehicle routing problems with Athos. Empir. Softw. Eng. 27, 180 (2022).
23. Challenger, M., Kardas, G., Tekinerdogan, B.: A systematic approach to evaluating domain-specific modeling language environments for multi-agent systems. Softw. Qual. J. 24, 755–795 (2016).
24. Sadeghi-Bazargani, H., Tabrizi, J.S., Azami-Aghdash, S.: Barriers to evidence-based medicine: a systematic review. J. Eval. Clin. Pract. 20, 793–802 (2014). https://doi.org/10.1111/jep.12222.
25. Wilk, S., Michalowski, M., Michalowski, W., Rosu, D., Carrier, M., KezadriHamiaz, M.: Comprehensive mitigation framework for concurrent application of multiple clinical practice guidelines - ScienceDirect. J. Biomed. Inform. 66, 52--71 (2017). https://doi.org/DOI:10.1016/j.jbi.2016.12.002.
26. Morris, J.E., Roderick, P.J., Harris, S., Yao, G., Crowe, S., Phillips, D., Duncan, P., Fraser, S.D.: Treatment burden for patients with multimorbidity: cross-sectional study with exploration of a single-item measure. Br. J. Gen. Pract. 71, e381–e390 (2021).
27. Meacham, S., Simon Fraser, Godwin Ukeje, Abiodun Adedeji: Customised computerised clinical protocol guidelines for medical education: the case of Chronic Kidney Disease (CKD) using domain-specific languages. Presented at the NET Conference 2024: Igniting change: breaking boundaries with digital innovation and creativity in healthcare education | Advance HE.
28. Palinkas, L., Horwitz, S., Green, C., Wisdom, J., Duan, N., Hoagwood, K.: Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. Adm. Policy Ment. Health Ment. Health Serv. Res. 42, 533--544 (2015). https://doi.org/DOI:10.1007/s10488-013-0528-y.
29. Malmqvist, J., Hellberg, K., Möllås, G., Rose, R., Shevlin, M.: Conducting the Pilot Study: A Neglected Part of the Research Process? Methodological Findings Supporting the Importance of Piloting in Qualitative Research Studies. Int. J. Qual. Methods. 18, 1609406919878341 (2019). https://doi.org/DOI:10.1177/1609406919878341.
30. Nielsen, J.: Usability Engineering. Academic Press (1994).
31. Green, T.R.G., Petre, M.: Usability Analysis of Visual Programming Environments: A 'Cognitive Dimensions' Framework. J. Vis. Lang. Comput. 7, 131--174 (1996). https://doi.org/DOI: 10.1006/jvlc.1996.0009.
32. Følstad, A.: Users' design feedback in usability evaluation: a literature review | Human-centric Computing and Information Sciences. Hum.-Centric Comput. Inf. Sci. 7, 19 (2017). https://doi.org/DOI: 10.1186/s13673-017-0100-y.
33. Hornbæk, K.: Current practice in measuring usability: Challenges to usability studies and research. Int. J. Hum.-Comput. Stud. 64, 79--102 (2006). https://doi.org/DOI: 10.1016/j.ijhcs.2005.06.002.