Measuring and Explaining Cognitive Load During Design Activities: A Fine-Grained Approach

Barbara Weber, Manuel Neurauter, Andrea Burattin, Jakob Pinggera and Christopher Davis

Abstract Recent advances in neuro-physiological measurements resulted in reliable and objective measures of Cognitive Load (CL), for example, using pupillary responses. However, continuous measurement of CL in software design activities, for example, conceptual modeling, has received little attention. In this paper, we present the progress of our work intended to close this gap by continuously measuring cognitive load during design activities. This work aims at advancing our understanding of WHEN and WHY designers face challenges. For this, we attempt to explore and explain the occurrence of CL using fine-granular units of analysis (e.g., type of subtasks, evolution of design artifact's quality, and manner of technology use). We expect implications for the future development of intelligent software systems, which are aware WHEN a particular designer experiences challenges, but also WHY challenges occur.

Keywords Business process management · Process modeling · Process model creation · Eye tracking · Cognitive load

1 Introduction

Contemporary software engineering practice differs fundamentally, as cloud-based apps and services do, from the monolithic mainframes of the 1980s. This presents a challenge, since timing and duration of software engineering design activities today is diffuse when compared to projects managed using structured methods and tools such as CASE (Computer Aided Software Engineering) and Integrated

B. Weber (✉)
Technical University of Denmark, Kongens Lyngby, Denmark
e-mail: bweb@dtu.dk

B. Weber · M. Neurauter · A. Burattin · J. Pinggera
University of Innsbruck, Innsbruck, Austria

C. Davis
University of South Florida, Tampa, USA

F.D. Davis et al. (eds.), *Information Systems and Neuroscience*,
Lecture Notes in Information Systems and Organisation 25,
https://doi.org/10.1007/978-3-319-67431-5_6

Development Environments (IDEs). It has become increasingly difficult to assess WHEN software engineers (designers in the following) experience challenges while conducting a design activity (e.g., creating a conceptual model or programming) and to explain WHY these challenges occur.

The cognitive demands imposed on the designer are commonly described as cognitive load (CL) [1]. Recent advances in neuro-physiological measurements resulted in reliable and objective measures of continuous CL [1]. However, continuous measurement of CL in software design activities up to now has received little attention. Moreover, there is no comprehensive understanding of the factors influencing a designers CL while conducting a software design activity.

Qualitative approaches in the context of software engineering typically identify challenges designers face through manual analysis and subsequent coding of data (e.g., [2, 3]) rather than the usage of neuro-physiological measurements of CL. Quantitative studies to CL measurements, in turn, are typically either perception-based and not continuous (e.g., using the NASA-TLX instrument for measuring CL [4]) or conducted as stimuli-response experiments where markers are available (i.e., it is known when the stimulus occurs) that can be related to the responses (i.e., change in pupil dilation), for example, [5]. In contrast to stimulus-response settings, where changes in cognitive load in response to a stimulus induced by experimenters are evaluated, investigating CL during design activities is less structured and inherent to how a designer's individual design process unfolds.

We intend to close this gap by measuring CL through objective, neuro-physiological measures in a more realistic work setting where no markers exist. We aim to make software engineering processes and their cognitive demands more tractable, advancing our understanding of WHEN and WHY designers face challenges. For this, we attempt to explore and explain the occurrence of CL using more granular units of analysis (including several process-oriented factors such as the type of sub-tasks; the evolution of the quality of the design artifact, and the manner of technology use) derived from the designer's interactions with the design platform and eye fixations on the various parts of the design platform.

This paper focuses on one frequently re-occurring software design activity, that is, conceptual modeling, but eventually aims at broadening the range of design components to include program blocks. Our research is expected to contribute towards a better understanding of WHEN and WHY high CL occurs in design activities by gathering empirical data regarding variations and changes in CL and various process-oriented factors to potentially explain these changes.

2 Cognitive Load in Design Activities

Design activities, for example, conceptual modeling, involve the construction of a mental model of the domain from an informal requirements description and its externalization using the elements provided by the modeling notation [6] by using the tools provided by a design platform, for example, the modeling editor [7]. During the externalization process, the designer evolves the design artifact, that is, conceptual model, through a series of interactions from an initial state through intermediate states to a final state reflecting the requirements of the domain. When performing a design activity, the designer exploits the malleability of their mental model to decompose cognitively 'digestible' sub-tasks, for example, a group of model elements. Recomposition maintains the integrity of the components and the intellectual control of the designer. This is the 'dance' of design that is choreographed using notations and design platforms [8]. The cognitive demands imposed on the designer are commonly described as CL, dependent on the task's inherent complexity, the design platform, the designer's domain knowledge, design expertise, and cognitive abilities. As a response to the cognitive challenges, the designer's CL changes throughout the course of the design activity [9]. To objectively measure CL and to determine WHEN designers are challenged, neuro-physiological tools, for example, eye tracking [1], can be used. For a particular designer we assume designer-specific factors to remain stable during a design activity. Possible explanations for changes in CL probably stem from changes in task difficulty throughout a design activity and the designer's interactions with design platform and design artifact. A messy intermediate design artifact could, for example, lead to higher CL when working on the artifact afterwards. To understand WHY challenges during the design process occur without having markers, we attempt to connect the CL data with data regarding the sub-tasks of the design process (to trace down differences in task difficulty throughout the design activity), the evolution of the design artifact and its quality, and the manner of technology use.

3 Research in Progress

Subsequently, we outline the current status of our work and provide details regarding our future endeavor.

3.1 Step 1: Data Collection

To continuously assess CL we measure pupil dilation, which (under conditions of controlled illumination) reliably indicates CL [1]. Alternative load measures and the reasoning for choosing pupil dilation for our study are discussed in [9]. Process-oriented factors as possible explanations for CL are measured by collating interactions with the design platform using Cheetah Experimental Platform (CEP) [10] and eye movement data (e.g., fixations) using the Tobii-TX300 eye tracker. For synchronizing interactions, fixations, and pupillary response data and for performing data treatment, we rely on a dedicated platform extending the capabilities of CEP towards analyzing CL [11, 12]. We collected data of 117 novice student modelers, who created a conceptual model using BPMN [9] after a training phase.

3.2 Step 2: Measuring Process-Oriented Factors

Step 2.1: Measure sub-task specific CL: For conceptual modeling, [13] showed the existence of the sub-tasks problem understanding, method finding, modeling, reconciliation, and validation. Since different types of sub-tasks involve different underlying cognitive processes, the changes in CL can stem from the type of sub-task the designer is currently engaged in. For this, we intend to automatically discover the different sub-tasks the user is engaged in at different periods of time when interacting with the design platform. For this, we rely on an existing task model [10, 13] and formulate the challenge of aligning the data coming from different modalities (i.e., neuro-physiological data, interactions with design platform) with the task model as a classification problem and solve it using supervised learning approaches, for example, Support Vector Machines [14]. For validation purposes the classifier will be compared with a previously defined gold standard. Initial results are promising [15, 16]. The alignment will be then used to slice the design activity into periods of time with start and end timestamps to calculate sub-task specific CL. Different measures such as the average CL, accumulated CL, or the instantaneous CL as suggested by Xie and Salvendy [17] can be calculated for the respective periods of time. For an overview of CL measurements in general and neuro-physiological measurements in particular please see [1].

Step 2.2: Measurement of the design artifact's quality evolution: When transforming a design artifact from one state into another, its quality (e.g., element alignment) can change and impact subsequent modifications. Put differently, a design artifact whose quality gradually degrades can make subsequent changes difficult by raising CL due to decreased readability. For this, quality will be operationalized as a set of properties [18, 19] (e.g., number of syntactical errors, alignment of elements). Values for each property can be calculated for each intermediate state. For this, we build upon infrastructure from the Austrian-funded ModErARe project [19].

Step 2.3: Conceptualize and measure manner of technology use: In addition, the manner of using the provided design platform can have an impact on CL, for example, tool features that are used effectively can lower CL, be ineffective, or even increase CL if used inappropriately. Here we plan to develop a rich conceptualization of technology use in line with [20] that goes beyond simple quantitative measures. Refactoring tools, for example, are frequently used as part of reconciliation sub-tasks with the goal to reduce CL of subsequent sub-tasks (e.g., by improving the understandability of the partial design artifact). When analyzing the potential impact of using refactoring tools, counting the number of its invocations is not sufficient, but it has to be considered whether its use was effective and led to improvements of the partial design artifact. Moreover, the potential benefit of refactoring depends on when in the process it is applied and how much the quality of the design artifact is impaired at the moment of its application. Therefore, our conceptualization of the manner of technology use will capture how a particular designer uses the design platform to accomplish a certain (sub-)task considering the state of the intermediate design artifact. Such data can be measured using the designer's interactions with the design platform.

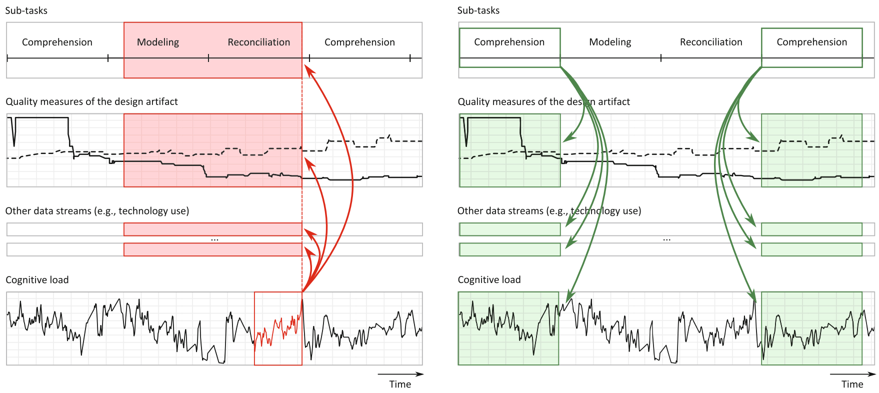
3.3 Step 3: Data Analysis

The analysis of the collected data can be grouped in two families: intra- and inter-subject analysis. The former case considers data stemming from different modalities, taking a single subject into account. The latter groups subjects by different aspects and analyzes the relationships among those groups. For intra-subject analysis, two different analyses might be considered (cf. Fig. 1). In both cases, the analysis combines the information streams we extracted, for example, the sub-tasks, the quality measurements of the designed artifacts, the manner of technology use, and the cognitive load.

After running through a data cleaning procedure described in [12] the starting point for the first analysis scenario (cf. Fig. 1a) is the identification of interesting time periods regarding CL. Then, we examine the immediate history of the other streams for identifying possible causes for the high CL. The second analysis approach (cf. Fig. 1b), on the contrary, slices the design activity into periods of time by considering process-oriented factors (i.e., sub-tasks, evolution of quality of design artifact, and manner of technology use). For example, as illustrated in Fig. 1, the design activity is sliced into sub-tasks. Another possibility of slicing the design activity could be identifying periods with high quality of the design artifact and periods with low quality. The analysis then compares periods of time of the same type (e.g., the different sub-tasks in Fig. 1b) in terms of differences in CL. Considering inter-subject analysis, subjects are grouped by process-oriented factors and then tested for group differences. For example, subjects could be grouped based on their modeling behavior (e.g., subjects with reconciliation phases just at the end of the modeling session versus subjects with reconciliation phases throughout the modeling session) and groups could be tested for differences in their average CL.

4 Summary and Expected Impact

In this paper, we investigate WHEN designers experience challenges by measuring CL and aim to explain changes in CL using different process-oriented factors. We have completed data collection and made substantial progress regarding the operationalization of process-oriented factors. Next, we plan to analyze the data as



(a) Starting from an interesting time period, analyze possible causes

(b) From specific modeling phases analyze quality of artifact and CL

Fig. 1 Approaches for the intra-subject data analysis

outlined. If the approach proves viable we intend to broaden our scope and to address other software design activities like programming.

Assuming we are able to demonstrate the impact of process-oriented factors on CL, we expect our research to result in revised guidelines on how to investigate design activities and evaluate design artifacts. Future research will be advised to depart from a pure black-box approach and to increasingly consider process-oriented factors impacting CL (or other antecedents of task performance). We further expect implications for the development of neuro-adaptive systems that are not only aware WHEN a particular designer experiences challenges, but also WHY and can react with personalized feedback or adaptation.

References

1. Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S.Z., Khawaji, A., Conway, D.: Eye-Based Measures, pp. 75–85. Springer International Publishing, Cham (2016)
2. Hungerford, B.C., Hevner, A.R., Collins, R.W.: Reviewing software diagrams: a cognitive study. IEEE Trans. Softw. Eng. **30**, 82–96 (2004)
3. Haisjackl, C., Zugal, S., Soffer, P., Hadar, I., Reichert, M., Pinggera, J., Weber, B.: Making sense of declarative process models: common strategies and typical pitfalls. In: Proceeding of Conference on BPMDS'13, pp. 2–17 (2013)
4. Marculescu, B., Poulding, S.M., Feldt, R., Petersen, K., Torkar, R.: Tester interactivity makes a difference in search-based software testing: a controlled experiment. Inf. Softw. Technol. **78**, 66–82 (2015)
5. Knapen, T., de Gee, J.W., Brascamp, J., Nuiten, S., Hoppenbrouwers, S., Theeuwes, J.: Cognitive and ocular factors jointly determine pupil responses under equiluminance. PLoS ONE **11**, 1–13 (2016)
6. Recker, J.C., Safrudin, N., Rosemann, M.: How novices design business processes. Inf. Syst. **37**, 557–573 (2012)

7. Soffer, P., Kaner, M., Wand, Y.: Towards understanding the process of process modeling: theoretical and empirical considerations. In: Proceedings of conference on ER-BPM'11, pp. 357–69 (2012)
8. Caputo, K.: CMM Implementation Guide: Choreographing Software Process Improvement (Unisys Series). Addison-Wesley, Boston (1998)
9. Neurauter, M., Pinggera, J., Martini, M., Burattin, A., Furtner, M., Sachse, P., Weber, B.: The influence of cognitive abilities and cognitive load on business process models and their creation. In: Proceedings of Conference on NeuroIS'15, pp. 107–115 (2015)
10. Pinggera, J., Zugal, S., Weber, B.: Investigating the Process of Process Modeling with Cheetah Experimental Platform. In: Proceedings of Conference on ER-POIS'10, pp. 13–18 (2010)
11. Weber, B., Neurauter, M., Pinggera, J., Zugal, S., Furtner, M., Martini, M., Sachse, P.: Measuring cognitive load during process model creation. In: Proceedings of Conference on NeuroIS'15, pp. 129–136 (2015)
12. Zugal, S., Pinggera, J., Neurauter, M., Maran, T., Weber, B.: Cheetah experimental platform web 1.0: cleaning pupillary data. Technical report (arXiv.org)
13. Pinggera, J.: The process of process modeling. PhD thesis, University of Innsbruck, Department of Computer Science (2014)
14. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**, 273–297 (1995)
15. Weber, B., Pinggera, J., Neurauter, M., Zugal, S., Martini, M., Furtner, M., Sachse, P., Schnitzer, D.: Fixation patterns during process model creation: initial steps toward neuro-adaptive process modeling environments. In: Proceedings of Conference on HICSS'16, pp. 600–609 (2016)
16. Burattin, A., Kaiser, M., Neurauter, M., Weber, B.: Eye tracking meets the process of process modeling: a visual analytic approach. In: Proceedings of Conference on TAProViz'16 (2016)
17. Xie, B., Salvendy, G.: Prediction of mental workload in single and multiple tasks environments. Int. J. Cogn. Ergonom. **4**, 213–242 (2000)
18. Burattin, A., Bernstein, V., Neurauter, M., Soffer, P., Weber, B.: Detection and quantification of flow consistency in business process models. Softw. Syst. Model., pp. 1–22 (2017)
19. Haisjackl, C., Burattin, A., Soffer, P., Weber, B.: Visualization of the evolution of layout metrics for business process models. In: Proceedings of Conference on TAProViz'16 (2016)
20. Burton-Jones, A., Straub, D.W.: Reconceptualizing system usage: an approach and empirical test. Inf. Syst. Res. **17**, 228–246 (2006)