

STUDY ON THE IMPACT OF COGNITIVE LOAD ON PERFORMANCE IN ENGINEERING DESIGN

**Zimmerer, Christoph;
Matthiesen, Sven**

Karlsruhe Institute of Technology

ABSTRACT

The objective evaluation of empirical studies is an important part when assessing demand and validating design methods. However, metrics that can also map cognitive processes during design are still lacking. In order to address this problem, an online study with 12 participants was conducted. The aim of this investigation was to find a relation between cognitive load and performance in engineering design tasks. To assess the cognitive load, the NASA-RTLX questionnaire was used as an established measurement tool and was related to the results achieved by the participants. The results show that there is a correlation between the two investigated parameters. Based on a statistical analysis a correlation between increasing cognitive load and a decrease in performance could be identified. The tasks used produce comparable results to other studies investigating cognitive load, but the task causing the highest cognitive load shows the widest scatter in performance. The u-curve as suggested by the state of the art was not visible in the study's results, but the cognitive load should be nevertheless used for studies of design processes, because it may reveal a need for methodical support.

Keywords: Design cognition, Human behaviour in design, Design engineering, cognitive load

Contact:

Zimmerer, Christoph
Karlsruhe Institute of Technology (KIT)
IPEK Institute of Product Engineering
Germany
christoph.zimmerer@kit.edu

Cite this article: Zimmerer, C., Matthiesen, S. (2021) 'Study on the Impact of Cognitive Load on Performance in Engineering Design', in *Proceedings of the International Conference on Engineering Design (ICED21)*, Gothenburg, Sweden, 16-20 August 2021. DOI:10.1017/pds.2021.537

1 INTRODUCTION

The objective analysis in empirical studies on the design process is necessary both in the assessment of demand and in the validation of design methods (Cash, 2018). Which research methods and proofs are necessary for successful validation remains a research gap in DRM (Gericke *et al.*, 2017) especially concerning objectivity of results. There is a lack of suitable, standardised methods, which allow an optimization of variables in experiments in design research (Üreten *et al.*, 2020). Dinar *et al.* (2015) describe that the need for clearly defined metrics to allow objective analysis is a general need in the design community. These metrics are necessary for the development of novel research methods that are also suitable for the objective evaluation of larger samples to perform more robust hypothesis testing. For this purpose, preferably methods with a quantitative approach should be used. (Dinar *et al.*, 2015) Especially in the analysis of cognitive activities in design, established research methods such as protocol analysis and interviews have limitations, since not all thought processes are verbalized and therefore are not detectable (Lloyd *et al.*, 1995). What is missing here are metrics, which can capture and measure the processes of design thinking in a wider field of application. A possible metric that meets these criteria is the cognitive load. Therefore, the presented study aims to investigate to what extent the cognitive load is suitable to assess the demand for design methods by quantifying cognitive processes.

1.1 Research methods to analyse design processes

For data collection in design research, a large number of different research methods exist. Ahmed (2007) gives an overview of different methods of classical empirical social research, which are used for design research. The choice of research methods depends on the construct to be investigated (e.g. creativity), which is made accessible by operationalising an observation or measurement (e.g. the number of ideas generated). The most frequently used methods are interviews, observations and protocol analysis (Ahmed, 2007).

However, with the above-mentioned research methods, the high effort involved in the evaluation limits the number of participants examined (Dinar *et al.*, 2015). According to Ahmed (2007) one hour of interview requires about nine hours for the transcription alone (Ahmed, 2007). Even if this has meanwhile become somewhat less due to specialized transcription software, the effort is still considerable. Typical studies in the journal “Design Studies”, which is considered the leading journal in the field of empirical studies on the design process, have fewer than 20 participants. Studies of the design process such as those conducted by Björklund (2013) (14 participants), Cramer-Petersen *et al.* (2019) (15 participants) or Nelius *et al.* (2020) (18 participants) are limited by the evaluation effort required concerning the number of participants. In studies with more participants (e.g. 33 participants in Kokotovich (2008)), it is usually not the design process itself but its result that is evaluated. Compared to other disciplines, the detailed analysis of procedures in the design process leads to smaller samples. In addition, the research methods used to analyse design processes in detail also influence the participants themselves. Concurrent think aloud, for example, can lead to additional cognitive demands during data acquisition due to the verbalisation of thoughts (van Someren *et al.*, 1994) and can thus lead to a corruption of the collected data. Due to such influences, the established research methods are not suitable for objective investigation of cognitive demands in the design process.

In summary, it can be said that there is already a broad variety of research methods to investigate the design process. However, there is still a lack of methods and metrics that can quantitatively measure the cognitive processes in design. In particular without increasing the cognitive demands of the participants.

1.2 Cognitive load

1.2.1 Approaches to assess cognitive load

Cognitive load theory originally derives from educational research. It assumes that learning involves a cognitive load (CL) that is limited by the capacity of the working memory. The term cognitive load includes several aspects like cognitive workload, mental strain and mental effort. All these terms are typically applied to describe the use of working memory. (Dias *et al.*, 2018) In this context, both problem solving and information processing play an important role. (Paas and van Merriënboer, 1994) Cognitive load for measuring cognitive demands is also widely used in other areas outside the field of educational research. According to Paas and van Merriënboer (1994), the cognitive load is a multi-dimensional parameter that represents the cognitive demands required to complete a task.

There are different approaches to measure cognitive load. One approach focuses on methods that use physiological signals such as heart rate variability (HRV) or pupillary response. A second approach builds on questionnaire-based survey methods, which are mainly used. A distinction is made between one-dimensional and multidimensional questionnaire-based methods. Hill *et al.* (1992) compare the four most widely used approaches:

The Modified Cooper-Harper scale (MCH) is a one-dimensional evaluation scale with 10 points that leads to a global assessment of the cognitive load. The rating scale uses a decision tree to help the participant to determine the most appropriate rating.

Another method is the Overall Workload Scale (OWS), which provides values on a one-dimensional scale from 0 for a low cognitive load to 100 for a high cognitive load.

The Subjective Workload Assessment Technique (SWAT) is a multidimensional survey technique that uses three dimensions to measure workload: time load, mental effort load and psychological stress load. Each of the three dimensions is in turn measured in three levels: low (1), medium (2) and high (3). From this assessment, a total workload is then calculated.

The fourth approach is the National Aeronautics and Space Administration Task Load Index (NASA-TLX) according to Hart and Staveland (1988). This approach uses a total of six dimensions to assess cognitive load: mental demand, physical demand, temporal demand, subjective performance, effort and frustration. Each of these dimensions is in turn assessed on a twenty-point scale. In the original NASA-TLX, these dimensions are then multiplied by weighting factors and added together to give a total value that indicates the overall cognitive load. These weighting factors are determined by pairwise comparison between all dimensions. However, it is also possible to omit the weighting factors and add the individual dimensions directly. This procedure is called NASA Raw TLX (NASA-RTLX). Whether this form of survey is more or less sensitive has not yet been conclusively clarified (Hart, 2006), so both are considered to be equivalent.

All four described approaches described above have already shown that they are able to assess cognitive load. In a comparison of the approaches Hill *et al.* (1992) came to the conclusion that the NASA-TLX and the OWS have the highest level of sensitivity and acceptance. In detail, the OWS is more suitable as a screening tool and NASA-TLX is more suitable for recording more detailed data. (Hill *et al.*, 1992) Therefore the NASA-TLX seems to be the most suitable for this investigation.

One challenge in the evaluation of NASA-TLX is the interpretation of the scores (Hart, 2006). For this reason, Grier (2015) analysed 237 surveys that used the NASA-TLX to measure cognitive workload. From the analysis of this large amount of data, typical ranges of cognitive load values for general and specific disciplines could be identified. Overall, the values ranged from 6 to 89. However, the majority of the data (80%) only ranged from 26 to 68. However, typical values were not separately identified for the field of engineering design in this investigation.

In summary, there are many established methods for measuring cognitive load from other areas, but the NASA-TLX questionnaire according to Hart and Staveland (1988) is the most suitable one. This is partly due to the high accuracy of the method, but also to the numerous applications in other domains. This allows the definition of typical ranges of cognitive load scores which can also be applied to the design process.

1.2.2 Cognitive load in studies of design processes

To use the cognitive load in design research, it is necessary to understand its effects on design processes. One impact of the cognitive load is on performance. The connection between these two variables is a curvilinear (or inverted-U) relationship, which was first described in the theory of Yerkes and Dodson (1908) and has already been shown in studies from other areas (Bruggen, 2015). The theory according to Yerkes and Dodson (1908) states that performance increases with increasing cognitive load up to a certain point. From this point on, overstraining occurs and causes a decrease in performance (see Figure 1). A low cognitive load results in a kind of understraining and therefore also in a decrease of performance.

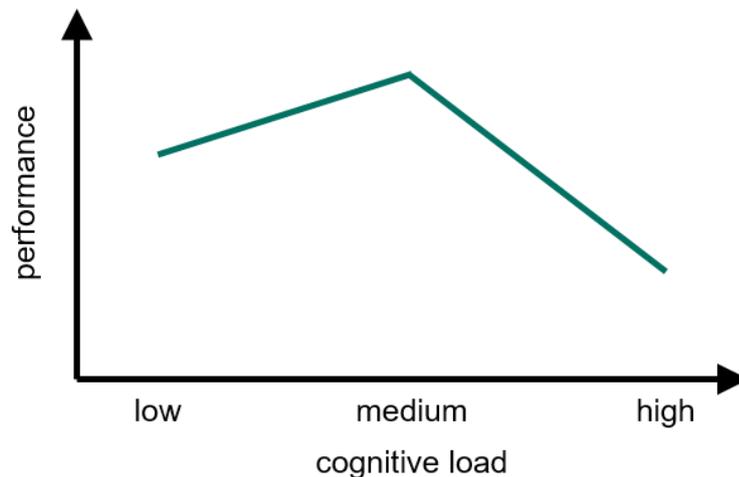


Figure 1. The inverted-U relation between cognitive load and performance following Bruggen (2015)

In the field of engineering design, there have been few studies that examine the cognitive load of designers (Howard *et al.*, 2008). These studies are mainly located in conceptual design to investigate the influences between cognitive demands and the creativity of solutions. (Sun and Yao, 2012; Nelson and Menold, 2020)

Sun and Yao (2012), for example, examined the cognitive efficiency of experienced and inexperienced designers during concept development. The cognitive efficiency was operationalized by the relation between mental effort and creativity. Howard *et al.* (2008) especially identify the potential to enhance creative performance by capturing and focusing cognitive capabilities. This is also intended to improve the quality of the solutions developed.

Cognitive load could be a very promising metric to study cognitive processes and their influence on performance. However, there is a lack of research in design that shows a connection between cognitive load and parameters of the design process, such as performance. Such a connection could be used to measure overstrain in design using the cognitive load and to reduce it in a targeted manner by applying the appropriate design methods.

1.3 Objectives

The aim of this study is to investigate the relationship between cognitive load and performance in engineering design tasks. For this purpose the following research question shall be answered:

RQ: Is there a relationship between cognitive load and performance in engineering design?

2 METHODS

The following chapter describes the study that was conducted to answer the research question. The structure of the study design and the evaluation procedure is described in detail.

A total of 15 participants took part in the study. A requirement for participation was that the participants were students in the fields of mechanical engineering or mechatronics and successfully completed a mechanical design course. This was to ensure that all participants had sufficient and comparable basic technical knowledge.

2.1 Tasks

Each participant was asked to work on six different tasks. The tasks spanned different fields of engineering design. A more detailed description of each task can be found in Table 1. In addition, there were two small introductory tasks at the beginning to familiarise the participants with the type of data acquisition. However, these introductory tasks were not included in the evaluation.

The different tasks are briefly described in Table 1:

Table 1. Description of the different tasks

task 1	analysis of technical drawing	In this task, participants should analyse a technical drawing of a gearbox, including identification of various design details such as the bearing concept and gear stages.
task 2	functional analysis 1: cartridge press	In this task a cartridge press represented by a technical drawing should be analysed. The aim of the task was to evaluate the influence of several parameters on the clamping effect of the reset mechanism.
task 3	functional analysis 2: snap-fit assembly	A snap-fit connection is a snap connection with a combination of non-positive and positive locking. Subject of the task was to assess the influence of various design parameters. In each case, the effect of an increase in the respective parameter on the retention force of the snap connection was to be evaluated.
task 4	considering manufacturing processes in design	This task deals with the manufacturing of a sheet metal bending part. For the unfolding of a given sheet metal part, the different bending steps should be considered in order to obtain the 3D part that is also given.
task 5	failure analysis	In this task various failure scenarios of bearings were to be assessed. For this purpose, the load on the shaft leading to the respective damage scenario should be identified for each given case.
task 6	load analysis in machine elements	In this case, a section of a technical drawing of a shaft was given. For several marked machine elements, the load that is applied to the component by clamping it with a shaft nut should be assessed here.

The purpose of the different tasks was to cover a similar range on the cognitive load scale as described in the results of [Grier \(2015\)](#). Both the total range and the range in which the majority of data sets (80%) are located were compared.

2.2 Data collection

The data collection was carried out through an online setup of the study. The university's own learning platform was used.

The basic structure of the study was designed in a way that after each task a query of the cognitive load via the NASA-RTLX was carried out. The questionnaire was based on the dimensions described by [Hart \(2006\)](#). Only the dimension physical demand was excluded, as the study was conducted solely on computer. Therefore no differences in physical load were expected between tasks. Previous research has also shown that the design process should be considered more as mental and not as physical activity ([Nikulin et al., 2019](#)).

The tasks were shortly introduced in both text and image form to the respective problem, followed by several questions to which the participants could enter their solution.

2.3 Data analysis

The data analysis was carried out in four consecutive steps:

1. First, the results from the NASA-RTLX questionnaires were evaluated. For each participant and task the values of the five individual dimensions were normalised to obtain the cognitive load on a scale from 0 to 100. Subsequently, the values were grouped according to the tasks.
2. In a second step, the tasks were evaluated in terms of performance. For this purpose, a defined number of different items were determined for each task in advance of the execution. In the analysis, the participants' answers for each item were evaluated for every task and classified as correct or incorrect with regard to the pre-defined solution. At the end of the performance evaluation, the quotient of correct items in relation to the total number of items available was calculated for each task. Thus, the performance for each task can be determined specifically in [%] and compared with the values from the other tasks.
3. The results from the first two steps were then combined to investigate possible relationships between the two parameters. For this purpose, the performance values of the individual tasks

were set in relation to the result of the respective NASA-RTLX questionnaire. This way, one data point, consisting of the two parameters cognitive load and associated performance, is created for each participant and task. The resulting data set is used to search for a correlation between the two parameters.

4. A statistical evaluation conducted in a fourth step completed the analysis. The SPSS Statistics 26 software by IBM was used for evaluation. The confidence interval for the determination of significant results was set at 95%.

Two tests were used for the actual statistical evaluation. The Spearman correlation test was used to search for correlations between the cognitive load and performance. This is to be used because the data was not normally distributed.

For the investigation of differences between the individual performance values or the cognitive load values, the Mann-Whitney U test was used, since normal distribution of data could not be assumed.

3 RESULTS

Of the 15 participants, only 12 have fully completed the study. Therefore, only these 12 were considered for further evaluation. In the following chapters, the results from the NASA-RTLX questionnaires and the processing of the tasks are presented individually and in relation to each other.

3.1 Cognitive load in the different tasks

In the first step the NASA-RTLX questionnaires were evaluated as described in chapter 2.3 and the results for the individual tasks were examined. Figure 2 gives an overview of the results from the evaluation of the NASA-RTLX questionnaires separately for each task.

Across all tasks, the cognitive load scatters in a range between 7 and 95, which means that the values are very comparable with the range of 6 to 89 determined by Grier (2015). Furthermore, the range in which the majority of the data (80%) is located is also very close to the values reported by Grier (2015). In this study, 80% of the data points are between 23 and 71 and the comparative area is between 26 and 68.

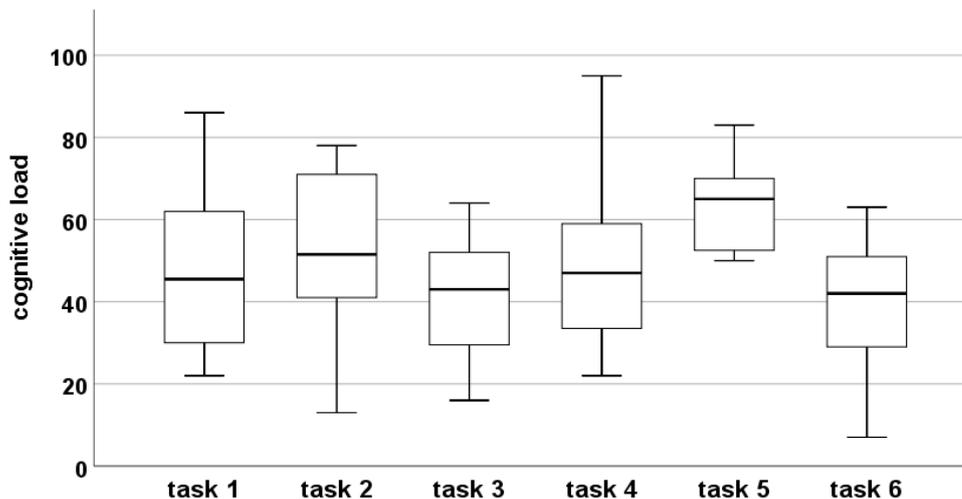


Figure 2. Cognitive load ratings for the different tasks

Looking at the tasks individually, it is visible (see Figure 1Figure 2) that they also produce different values in the cognitive load. Task 5 (failure analysis) is by far the most difficult task for the participants with the lowest scatter (median = 65; SD = 10.2). The remaining tasks are relatively close according to the median cognitive load but differ in their scatter. Task 3 (functional analysis; median = 43.0; SD = 15.8) and task 6 (load analysis in machine elements; median = 42.0; SD = 16.9) have a lower scatter than task 1 (analysis of technical drawing; median = 45.5; SD = 20.3), task 2 (functional analysis; median = 51.5; SD = 19.4), and task 4 (considering manufacturing processes in design; median = 47.0; SD = 21.3).

Task 5 (failure analysis) also produces statistically significantly higher values in the cognitive load than task 1 (analysis of technical drawing; two-tailed Mann-Whitney U test, $U = 36$, $p = 0.037$), task 3 (functional analysis; two-tailed Mann-Whitney U test, $U = 16$, $p = 0.001$), task 4 (considering manufacturing processes in design; two-tailed Mann-Whitney U test, $U = 34$, $p = 0.028$), and task 6 (load analysis in machine elements; two-tailed Mann-Whitney U test, $U = 13.5$, $p = 0.001$).

3.2 Performance in the different tasks

Following the cognitive load, the performance of the participants in the different tasks was evaluated as a second step. Figure 3 provides a closer look at the performance of the participants in the different tasks:

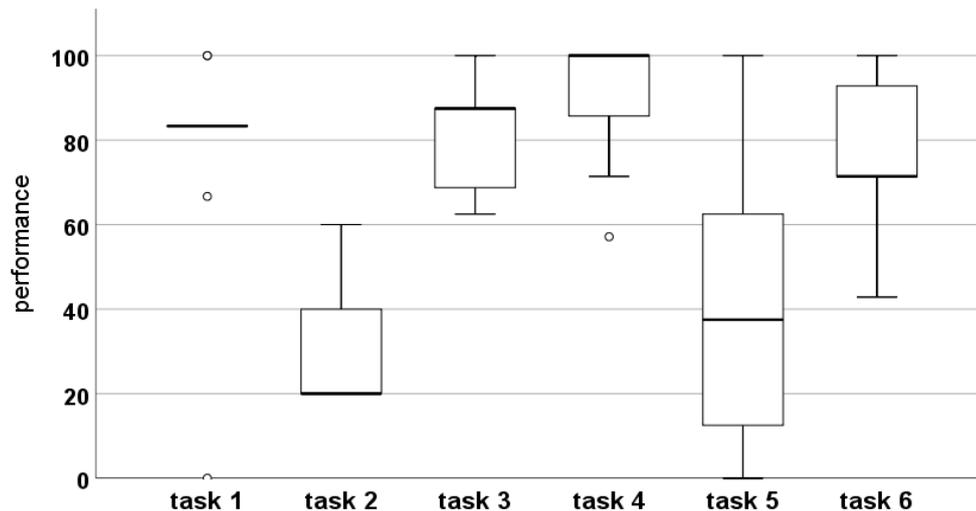


Figure 3. Results of the performance within the different tasks

It can be seen that the participants performed considerably better in task 1 (analysis of technical drawing; median = 83.3; SD = 26.0), task 3 (functional analysis; median = 87.5; SD = 13.6), task 4 (considering manufacturing processes in design; median = 100; SD = 14.2) and task 6 (load analysis in machine elements; median = 71.4; SD = 19.4) than in task 2 (functional analysis; median = 20.0; SD = 15.9) and task 5 (failure analysis; median = 37.5; SD = 32.8).

The scatter in the individual tasks is rather low with values between 14.2 and 26.0 with the exception of task 5 (failure analysis) where the scatter is significantly higher with 32.8.

3.3 Correlation between cognitive load and performance

Combining the results from the two datasets cognitive load and Performance results in the scatter plot shown in Figure 4. Each data point represents one task processed by one participant. A correlation analysis was conducted to determine whether there is a correlation between the two parameters. The result is shown as a linear line in the figure. As the data is not normally distributed, a correlation analysis according to Spearman is used. The cognitive load indicated by the participants correlates significantly with the performance achieved in the corresponding tasks, $r = -0.243$, $p = 0.040$, $n = 72$. According to Cohen (1992) this value is attributed to a weak effect.

A correlation analysis based on the assumption of a quadratic relationship similar to the curvilinear (or inverted-U) relationship did not yield a convincing result.

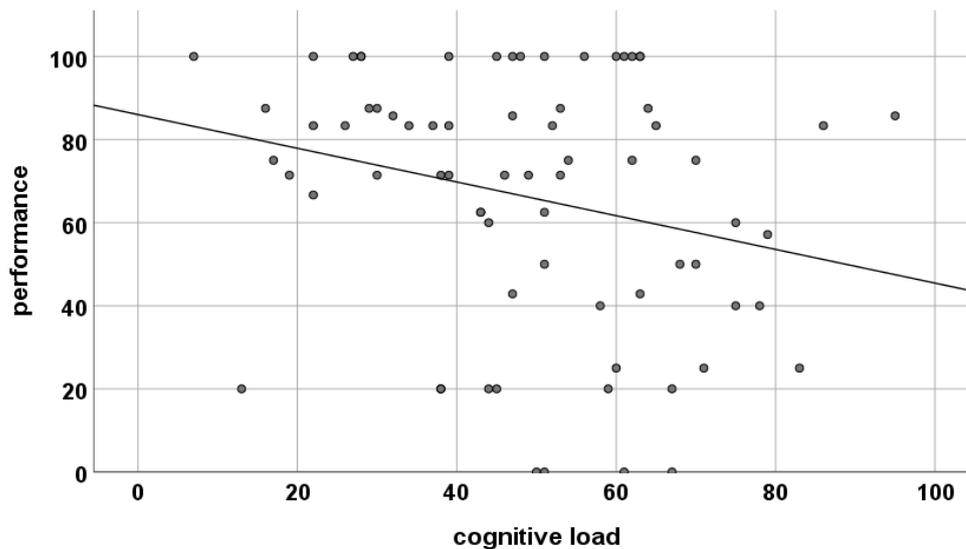


Figure 4. Correlation between cognitive load values and performance across all tasks and participants displayed as a negative trending line

4 DISCUSSION

The aim of this research was to investigate the relationship between cognitive load and performance in design. A general decrease in performance with increasing cognitive load could be detected. This performance decrease can be interpreted as cognitive overstraining caused by difficulties in processing the given tasks. However, an additional drop in performance at low levels of cognitive load could not be shown. Thus, this study has been unable to demonstrate that there is a specific level of cognitive load at which the performance reaches an optimum.

4.1 Suitability of the tasks

Since there have been very few investigations of the cognitive load in the field of engineering design so far and these are limited to conceptual design, the suitability of the applied study design must be evaluated. For this purpose, the cognitive load values are compared with common values from other disciplines, which were investigated and summarised by Grier (2015). This comparison showed that the tasks used in this study achieved very similar values to those obtained in investigations from other areas. Altogether, the six different tasks covered a wide range on the cognitive load scale. The distribution of the values is generally very close to the values that were identified in the investigation of studies from other areas which were using the NASA-RTLX (Grier, 2015). This leads to the conclusion that the tasks as a whole are well suited to investigate the cognitive load in engineering design.

One imitation of the tasks used, however, is that there is very little difference between the different tasks. As can be seen from Figure 2, the only task that seemed to be significantly more difficult than the others was task 5 (failure analysis), which makes it hard to derive generally statements on the context investigated. However, differences can be seen in the distribution of the individual tasks. Especially tasks 1, 2 and 4 show a strong scattering of cognitive load values. From this can be derived that the individual difficulties of the participants in certain areas are more important than a difficulty induced by the individual task.

When looking at the performance in the individual tasks, a contradictory situation emerges. While the tasks that scattered very much in the cognitive load show a very low degree of scatter in the performance. Task 5 (failure analysis), which has the highest values with the lowest scatter in the cognitive load, scatters the most in performance. This suggests that a high cognitive load does not necessarily result in poor performance.

4.2 Correlation between cognitive load and performance

When the two data sets from cognitive load and performance are combined in a single diagram (Figure 4), it is difficult to establish a relationship between the two variables based on the individual data

points. Although a correlation analysis has shown a significantly decreasing correlation, it is rather weak. This is also illustrated by the weak effect according to Cohen (1992). From this, it can be concluded that although the available data provide a weak indication of the right part of the inverted-U curve according to Yerkes and Dodson (1908) and Bruggen (2015). This effectively represents the area of overstrain. The left part of the inverted-U curve, which stands for understrain, could not be mapped. A possible explanation for this situation is that the tasks used can be described as rather “simple” based on the NASA-RTLX used, but were not so trivial that an actual understraining occurred.

4.3 Limitations

The effect of the statistical outcome is limited by the relatively small number of participants in this study ($n = 12$). Furthermore, due to the decision for an online setup for the study, it was not possible to fully control the study conditions during the processing. This may have led to deviating influences on the individual dimensions of the NASA-RTLX.

5 CONCLUSION

The identified correlation between cognitive load and performance has the potential to be useful for future investigations in design research. Of particular interest is the area of overstrain, as this is seen as having great potential for design methods that could increase performance in design by reducing overstrain.

In the study presented here, only a weak correlation between cognitive stress and performance was found, but this is partly due to the relatively small number of subjects ($n = 12$) and the fact that the study was conducted using an online platform. In order to use cognitive load more effectively in design research, the correlation between cognitive load and performance needs to be better understood. Further research is needed in this area. In particular, the precision of the measurement of cognitive stress can be improved by recording physiological signals like heart rate variability (HRV) or pupillary response in addition to the questionnaires used. If it turns out that there is a strong correlation, cognitive stress could be used as a metric in a research method for requirements assessment and validation of methods in design.

REFERENCES

- Ahmed, S. (2007), “Empirical research in engineering practice”, *J. of Design Research*, Vol. 6 No. 3, pp. 359–380. 10.1504/JDR.2007.016389.
- Björklund, T.A. (2013), “Initial mental representations of design problems: Differences between experts and novices”, *Design Studies*, Vol. 34 No. 2, pp. 135–160. 10.1016/j.destud.2012.08.005.
- Bruggen, A. (2015), “An empirical investigation of the relationship between workload and performance”, *Management Decision*, Vol. 53 No. 10, pp. 2377–2389. 10.1108/MD-02-2015-0063.
- Cash, P.J. (2018), “Developing theory-driven design research”, *Design Studies*, Vol. 56, pp. 84–119. 10.1016/j.destud.2018.03.002.
- Cohen, J. (1992), “A Power Primer”, *Psychological Bulletin*, Vol. 112 No. 1, pp. 155–159.
- Cramer-Petersen, C.L., Christensen, B.T. and Ahmed-Kristensen, S. (2019), “Empirically analysing design reasoning patterns: Abductive-deductive reasoning patterns dominate design idea generation”, *Design Studies*, Vol. 60, pp. 39–70. 10.1016/j.destud.2018.10.001.
- Dias, R.D., Ngo-Howard, M.C., Boskovski, M.T., Zenati, M.A. and Yule, S.J. (2018), “Systematic review of measurement tools to assess surgeons’ intraoperative cognitive workload”, *The British journal of surgery*, Vol. 105 No. 5, pp. 491–501. 10.1002/bjs.10795.
- Dinar, M., Shah, J.J., Cagan, J., Leifer, L., Linsey, J., Smith, S.M. and Hernandez, N.V. (2015), “Empirical Studies of Designer Thinking: Past, Present, and Future”, *Journal of Mechanical Design*, Vol. 137 No. 2, p. 247. 10.1115/1.4029025.
- Gericke, K., Eckert, C. and Stacey, M. (2017), “What do we need to say about a design method?”, *21th International Conference on Engineering Design (ICED 2015)*.
- Grier, R.A. (2015), “How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores”, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 59 No. 1, pp. 1727–1731. 10.1177/1541931215591373.
- Hart, S.G. (2006), “Nasa-Task Load Index (NASA-TLX); 20 Years Later”, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50 No. 9, pp. 904–908. 10.1177/154193120605000909.

- Hart, S.G. and Staveland, L.E. (1988), “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research”, *Advances in Psychology*, Vol. 52, pp. 139–183. 10.1016/S0166-4115(08)62386-9.
- Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner, A.C., Zaklad, A.L. and Christ, R.E. (1992), “Comparison of Four Subjective Workload Rating Scales”, *Human Factors*, Vol. 34 No. 4, pp. 429–439.
- Howard, T.J., Culley, S.J. and Dekoninck, E. (2008), “Describing the creative design process by the integration of engineering design and cognitive psychology literature”, *Design Studies*, Vol. 29 No. 2, pp. 160–180. 10.1016/j.destud.2008.01.001.
- Kokotovich, V. (2008), “Problem analysis and thinking tools: an empirical study of non-hierarchical mind mapping”, *Design Studies*, Vol. 29 No. 1, pp. 49–69. 10.1016/j.destud.2007.09.001.
- Lloyd, P., Lawson, B. and Scott, P. (1995), “Can concurrent verbalization reveal design cognition”, *Design Studies*, Vol. 16 No. 2, pp. 237–259.
- Nelius, T., Doellken, M., Zimmerer, C. and Matthiesen, S. (2020), “The impact of confirmation bias on reasoning and visual attention during analysis in engineering design: An eye tracking study”, *Design Studies*, Vol. 71. 10.1016/j.destud.2020.100963.
- Nelson, J. and Menold, J. (2020), “Opening the black box: Developing metrics to assess the cognitive processes of prototyping”, *Design Studies*, Vol. 70, p. 100964. 10.1016/j.destud.2020.100964.
- Nikulin, C., Lopez, G., Piñonez, E., Gonzalez, L. and Zapata, P. (2019), “NASA-TLX for predictability and measurability of instructional design models: case study in design methods”, *Educational Technology Research and Development*, Vol. 67 No. 2, pp. 467–493. 10.1007/s11423-019-09657-4.
- Paas, F.G. and van Merriënboer, J.J. (1994), “Instructional control of cognitive load in the training of complex cognitive tasks”, *Educational Psychology Review*, Vol. 6 No. 4, pp. 351–371.
- Sun, G. and Yao, S. (2012), “Investigating the relation between cognitive load and creativity in the conceptual design process”, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 56 No. 1, pp. 308–312.
- Üreten, S., Eisenmann, M., Nelius, T., Garrelts, E., Krause, D. and Matthiesen, S. (2020), “Current challenges and solution approaches in empirical engineering design research – a workshop for empirical research”, *Proceedings of the Design Society: DESIGN Conference*, Vol. 1, pp. 61–70. 10.1017/dsd.2020.280.
- van Someren, M.W., Barnard, Y.F. and Sandberg, J.A.C. (1994), *The think aloud method: A practical guide to modelling cognitive processes*, Academic Press, London.
- Yerkes, R.M. and Dodson, J.D. (1908), “The relation of strength of stimulus to rapidity of habit-formation”, *Punishment: Issues and Experiments*.