IDENTIFYING SUCCESSFUL APPROACHES DURING TESTING ACTIVITIES IN ENGINEERING DESIGN

Liewerenz, Oliver;
Grauberger, Patric;
Nelius, Thomas;
Matthiesen, Sven

Karlsruhe Institute of Technology

ABSTRACT

Testing activities to gain specific design knowledge play an essential role in engineering design, when a structure needs to be developed, for which knowledge from analytical models or documentation is missing. As research into these testing activities is difficult, few insights into successful approaches exist.

In this contribution, we investigate testing activities to gain specific design knowledge through a laboratory task, where 10 engineering students optimize a system using a web-based process chain including rapid prototyping and testing. Design and testing data are acquired from 110 prototypes in 3 hours. A differentiation of performance is conducted and approaches of high- and low-performers are investigated to identify patterns.

Based on these patterns, hypotheses, and metrics indicating successful and non-successful approaches are derived as basis for development of metrics for testing to gain specific design knowledge. A successful approach was overstep the limit, where participants accept destruction of their system to identify boundaries. An unsuccessful approach was the change of many parameters in later tests. These hypotheses and their metrics can then be used in development and validation of support for testing.

Keywords: Data-driven engineering design, Testing activities, Design methods, Embodiment design, Process modelling

Contact:
Liewerenz, Oliver
Karlsruhe Institute of Technology
Germany
oliver.liewerenz@kit.edu

Cite this article: Liewerenz, O., Grauberger, P., Nelius, T., Matthiesen, S. (2023) 'Identifying Successful Approaches during Testing Activities in Engineering Design', in *Proceedings of the International Conference on Engineering Design (ICED23),* Bordeaux, France, 24-28 July 2023. DOI:10.1017/pds.2023.221

1 INTRODUCTION

In product development processes, prototyping and especially testing activities are drivers of costs and time needed to develop a product. Testing in general is "the process of operating a system or component under specified conditions, observing or recording the results, and making an evaluation of some aspect of the system or component" (IEEE, 1990). Testing activities can be differentiated into planned and unplanned activities. Many of these activities can be defined in project planning and serve purposes like ensuring product functionality or fulfilment of legal requirements (Engel, 2010). During product development, many different testing activities can occur (Tahera et al., 2019).

A case, where unplanned testing activities emerge, is caused by missing specific design knowledge (Hubka and Eder, 1990), which cannot be gained through enquiry, simulation, or consulting of experienced engineering designers (Matthiesen, 2021). This is also a cause for iterations, in which problem and solution evolve together in cycles (Dorst and Cross, 2001; Clarkson et al., 2007; Meboldt et al., 2013).

At the start of a development project, it is mostly unclear, which parts of the specific design knowledge are missing and need to be gained through testing. Testing activities to gain specific design knowledge can be described using the Function-Behaviour-Structure (FBS) ontology (Gero and Kannengiesser, 2004). Tests with physical structures (e.g. prototypes) are conducted to identify the behaviour derived from structure (Bs), which then can be compared to the expected behaviour (Be), enabling insights into the relations of structure and system behaviour. These insights enable design engineers to synthesize a suitable product structure for the desired behaviour. As success in these testing activities contributes to overall success of product development, this topic is of interest to design research.

Studies on testing activities are often conducted using examples of development projects in industry. The investigated cases take place over a long time and need a great amount of resources (Shabi et al., 2017). They are often evaluated through qualitative data acquisition like interview studies (Tahera et al., 2019). In retrospective interviews, testing activities to gain specific design knowledge are often missing, as participants seldom recall all the testing activities they have done, especially not the seemingly unsuccessful ones (Cash et al., 2016). The use of small numbers of participants over a long period of time complicates the transferability of results through many disturbing variables, low statistical power and possibly inflated effect size (Erichsen et al., 2019). In addition, different aims under various boundary conditions hamper comparability.

Laboratory environments enable increased comparability (Kroll and Weisbrod, 2020). Studies in the lab regarding testing activities also exist. They also show some difficulties, for example the time-consuming prototyping and difficult data acquisition. During the relatively long development processes, an unknown amount of testing activities occurs and detailed insights into them is not possible by investigating e.g. the stages of the prototypes (Barhoush et al., 2019).

Study environments that address the comparability through many similar tasks, which are conducted in parallel project based learning environments (Batliner et al., 2022), however the time consuming aspects remain (Grauberger et al., 2019). The problem is, that insights regarding testing activities to gain specific design knowledge are difficult to research, as testing activities often take place over a long period and are therefore difficult to capture. Due to the mentioned difficulties in studying testing, it is not yet known which testing activities are particularly successful in gain of specific design knowledge.

The research question derived from this challenge is as follows: *What are successful approaches for testing activities to gain specific design knowledge?*

To answer this question, we conducted a laboratory study to capture testing approaches comparably. In the study, engineering students performed a condensed embodiment design task in which a physical product must be optimized to achieve the expected behavior.

2 MATERIALS AND METHODS

In this contribution, we answer the research question through investigation of testing activities in a laboratory environment, where engineering students conduct a task in which specific design knowledge needs to be built up. A snap-fit connection as physical product needs to be optimized to reach its expected behaviour. In this task, testing activities can be performed in a short period through fast, iterative design and production cycles. Data acquisition is focussed on quantitative data and based on generated design data without interfering through qualitative data collection methods or relying on subsequent methods like questionnaires.

To investigate successful approaches in testing activities in a laboratory environment, certain boundary conditions have to be considered: First, the task should be close to real testing activities, as successful approaches identified in strongly deviating tasks are difficult to transform into insights for testing activities in practice. Second, to gain clear results, a laboratory study has to have a limited, uninterrupted time span of a few hours, as participants lose concentration and motivation over time and the influence of disturbance variables (e.g., when participants talk to each other after a session) increases by including breaks. The expenditure of larger studies (over multiple days) is also not to be neglected. The system, and also the process chain of modelling, manufacturing, and testing need to be manageable by the participants in the limited time span.

2.1 Task - snap-fit connection

A snap-fit connection is a broadly applied technical system, which has no standardisation. Even though some references exist on how to design them, due to the chosen material, no sample solution for the task is present in literature. It is therefore essential to conduct testing activities to gain the specific design knowledge needed to optimize the main function of reaching a required holding force. The chosen variant of a snap-fit connection in this study is made from HDF (high-density fibre board) by using a laser cutter. A snap-fit connection made from HDF can be produced in less than 20 seconds, whereas it would take well over 30 minutes to produce a 3D printed prototype. In addition, the accuracy of the laser cutter is a few hundredths of a millimetre, which is much better than 3D printing. For the design task, a HTML front end for a web-based computer-aided design (CAD) environment (Rhino® 7 Grasshopper model hosted by Shapediver GmbH), where parameters can be adapted by moving sliders, is applied. This accelerates the task of changing the design and preparing it for manufacturing compared to the original CAD tool. A live-rendered CAD model enables real-time visual feedback for the participants regarding their design decision. Figure 1 shows the snap-fit connection in the web-based CAD environment (left side) and the resulting HDF prototype (right side) with an enlarged picture of a design detail. The main function of a snap-fit connection is to release connected components in a non-destructively way at a defined holding force threshold. The geometry of the snap-fit hook influences the holding force. The design of the latching hook can be freely defined via 16 design parameters. By varying the design parameters, it is possible to enable a range of force from 0N (instant release) to release only through destruction of the hooks. The counterpart cannot be changed in this study. It is also the more robust part of the system, ensuring destruction of the hooks for unsuitable design choices.

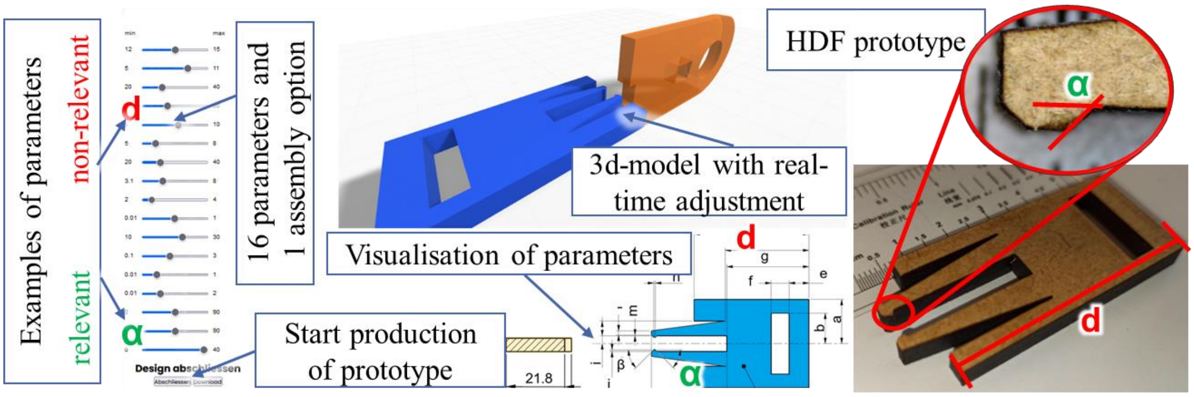

Figure 1. snap-fit connection with CAD model and HDF prototype

From the 16 parameters that can be adapted in the web-based CAD, eight have an influence on the holding force. An example is the angle α (top right), that influences the force distribution during holding. The other eight parameters are necessary for the geometry, however, in the chosen boundaries, they have no influence on the holding force. Here an example is the length d (bottom right), that influences only the mounting guidance. Further characterization of effect strength of the parameters is possible; however, for the aim of identifying approaches to testing, the binary granularity in influencing and non-influencing parameters is sufficient.

2.2 Testing in the lab - investigation environment

The task was to optimize a given snap-fit connection with a maximal holding force of ~10N (Bs) to a new maximal holding force of >200N (Be). This is the force a fictional competitor is able to reach. A boundary condition is the possibility to reuse the snap-fit connection four times without destroying it to ensure, that the task cannot be completed through a locking geometry with release through destruction, which would be in contradiction with the main function.

To enable as many testing activities as possible in the short time span, a low-threshold process chain is used. It consists of the web-based CAD, a cloud-based interface for rapid manufacturing, and a simple force-measuring device. This process chain is depicted in Figure 2.

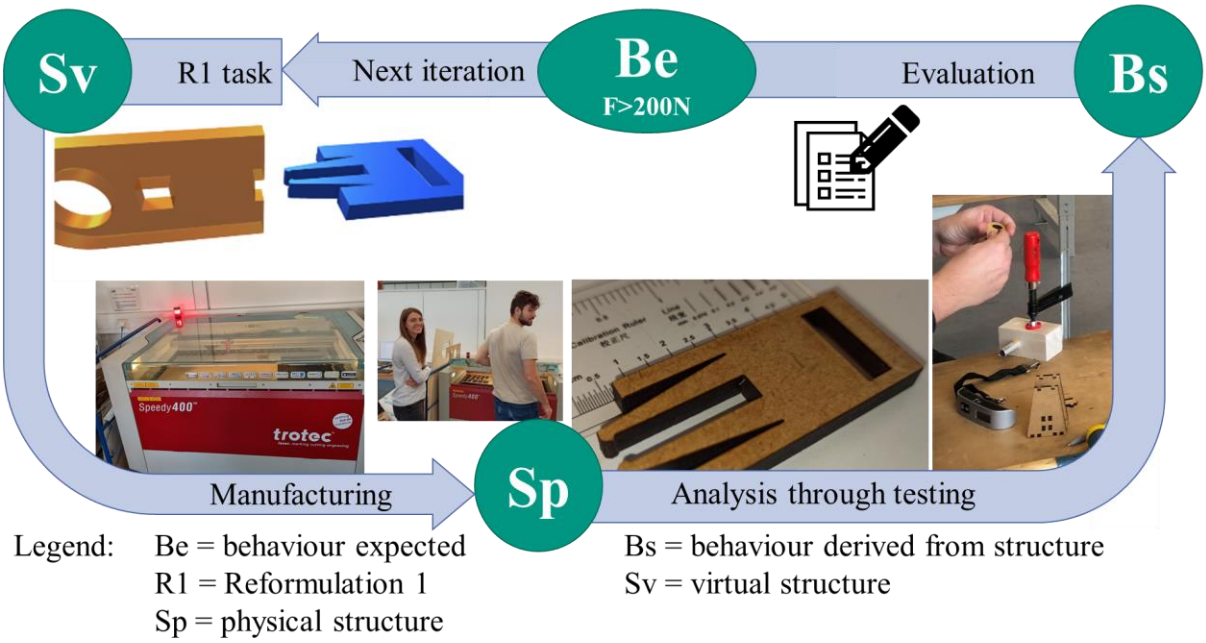

Figure 2. process chain for the testing activities to gain specific design knowledge

The snap-fit connection is designed in the web-based CAD model, which can be adapted in real-time using a server-based application (Shapediver). It is a *reformulation 1* (R1) activity from FBS, where changes in design parameters or their ranges of values are addressed, if the Bs is evaluated to be unsatisfactory (Gero and Kannengiesser, 2004). No new concepts are to be developed here and the result is a virtual structure.

From the web-based CAD application, a vector graphic is derived and transferred into a cloud, where it queues for manufacturing using a laser cutting machine. A standard snap-fit connection needs a production time of 17s. The counterparts of the connection have been manufactured before the study. Due to its precision of ±0.04mm, the designed snap-fit connections (virtual structure) can be produced with high reproducibility regarding Bs. Then the analysis through testing takes place, that derives the Bs from the physical structure (Gero and Kannengiesser, 2004). The snap-fit connection is mounted into a testing station and its maximal holding force (Bs) is measured by using a simple force measuring device. Then, the measured holding force (Bs) is compared to the Be in the evaluation and as the aim is to reach a force as high as possible, the next iteration starts.

In a preparation test of the study environment, this process chains enabled investigating testing activities in about 12 minutes per iteration (containing 8 minutes of thinking and changing the design, manufacturing, and testing), consequently about 15 iterations are possible in the period of 3h. For the study, 2 laser cutting machines with operators and 7 testing stations are prepared for the participants.

https://doi.org/10.1017/pds.2023.221 Published online by Cambridge University Press

2.3 Study procedure

A total of 10 participants took part in the study. All 10 were students of mechanical engineering at the Karlsruhe Institute of Technology (KIT). A requirement for students to participate was that they had completed at least the fourth semester of their studies and had also successfully passed the mechanical design and technical mechanics course, so that all participants had sufficient basic and comparable knowledge to complete the task.

As the introduction of the task and explanation of the process chain takes about 18 minutes, the participants have 2 hours and 42 minutes to work individually on optimizing the holding force of the snap-fit connection. During this time, no exchange between participants is allowed. After the introduction, participants will receive no further guidance from the study leader for the rest of the study. Then, the participants can begin the task on their personal notebooks. A QR code allows participants to access the prepared design environment (shown in Figure 2, top left). This QR code and a second QR code for saving the files for manufacturing will be provided to the participants throughout the study session. Each participant is provided with a sheet of notepaper, e.g., for sketches, calculations, and documentation of results. After three hours, the study ends and no further R1 activity is allowed.

2.4 Data collection

To investigate the participants' approaches, the parameter set of each iteration is captured automatically. It consists of the 16 design parameters and 9 meta data. The meta data include the randomly assigned participant identification, the iteration identification, a time stamp, the measured holding force of the reuse boundary condition (four repetitions) and an assembly parameter. When the snap-fit connection is broken during mounting, the value assigned for Bs is 0.

2.5 Data analysis

In the first step, the collected data of all participants is merged into a spreadsheet. For this purpose, the anonymized participant identification and iteration identification are used. The following parameters lay the basis for further analysis:

- Timestamps of iterations during the study
- Bs (resulting holding force)
- Design configuration (parameters and values, changes in iterations) - structure
- Percentage of function-relevant parameters in the changed parameters in an iteration

Overall success is determined through the maximal holding force achieved in the conducted tests. From this data, participants are differentiated into high- and low performers according to the maximal holding force reached.

- High performers: exceed the competitor holding force by >10% (>220N)
- Low performers: Do not reach the competitor holding force (<200N).
-

Participants in the gap between 200-220N are not evaluated in detail, as a clear differentiation of performance is deemed necessary to identify successful or unsuccessful approaches. Patterns in design parameters are then identified by comparing the changes of design parameters and resulting holding forces from high- and low performers in a visualisation of the data. Identified patterns are matched with the parameter dataset, the amount of changed parameters and percentage of function-relevant parameters of the participants. It is assessed by the authors, whether it is a single case or if this pattern is recurring in the dataset. The patterns lay the basis for derivation of hypotheses regarding key characteristics for successful approaches to testing activities to gain specific design knowledge.

3 RESULTS

In the following section, the data gained in the study are presented. At first, the raw data are shown in overview. Then, excerpts of the data visualisation are shown, at which approaches of the participants are described in detail.

https://doi.org/10.1017/pds.2023.221 Published online by Cambridge University Press

3.1 Raw data and overview

The data sets of the 10 participants were analysed. An overview is given in table 1.

Table 1. Overview of the evaluated parameters in the study

General		Force		Time		Design parameters		
P	IC	F_max [N]	F_first [N]	T_first [min]	T_mean [min]	Par_max	Par_med	rel_param [%]
P 01	13	203	19	5	11	9	2	50
P 02	15	284	75	7	10	9	2	74
P 03	8	250	38	5	16	8	2	75
P 04	14	131	61	1	10	7	2	65
P 05	12	143	0	2	11	11	4	47
P 06	8	210	28	5	16	9	3.5	72
P 07	11	235	56	3	13	6	1	98
P 08	11	162	0	0	13	17	4	52
P 09	9	240	23	6	16	7	3	82
P 10	9	245	46	0	13	9	3	82
Mean	11	210.3	34.6	3.4	12.88	9.2	2.65	68.5

Legend: P = participant; IC = iteration count; F_max = maximal holding force; F_first = force in first test; T_first = time to first test; T_mean = mean time between tests; Par_max = maximal number of changed parameters in an iteration; par_med = median of changed parameters in an iteration; rel_param = percentage of function-relevant parameters over all iterations

The participants performed 110 prototypes in total, resulting in 11 iterations on average (SD 2.5) per participant. The measured holding force ranged from an average of 34.6 N (SD = 25.26) in the first iteration to an average of 187.5 N (SD = 51.85) in the last iteration. The maximum holding force achieved by the participants (not necessary reached in the last iteration) averaged 210.3 N (SD = 50.45). A total maximum of 285 N could be achieved. An iteration took an average of 12.88 minutes. On average, 2.65 of the 16 design parameters were changed per iteration.

3.2 Categorization for uncovering patterns

To identify successful approaches based on patterns observed in the acquired data, we classify the participants into high-, and low-performers according to Section 2.5. Table 2 gives an overview of the average values reached for each of the categorized groups.

Table 2. Comparison of high- and low-performers

Performance	av. Iterations	F_first	F_max	T_first	T_mean	Par_mean	rel_param
Low	12.33	20.3N	145.33N	1 min	12 min	4	54.7 %
High	13.00	47.2N	250.80N	4 min	13 min	2	82.2 %

Legend: av: = average; F_first = force in first test; F_max = maximal holding force; T_first = time to first test; T_mean = mean time between tests; par_med = median of changed parameters in an iteration; rel_param = percentage of function-relevant parameters over all iterations

The categorization into high- and low-performers enables more specific investigation of the data described in Section 3.2. The difference regarding the maximal holding force indicates a clear differentiation of these categories. To identify patterns, the authors sifted the diagrams and searched for characteristics in the data, which were then discussed and either included or discarded. Due to this qualitative assessment, the patterns are seen as data-based description of approaches in testing activities. Patterns occurring in the iterations are visualised in diagrams of holding force over time. The percentage of function-relevant parameters is also noted in each iteration. Figure 3 shows the excerpt of the low-performers.

https://doi.org/10.1017/pds.2023.221 Published online by Cambridge University Press

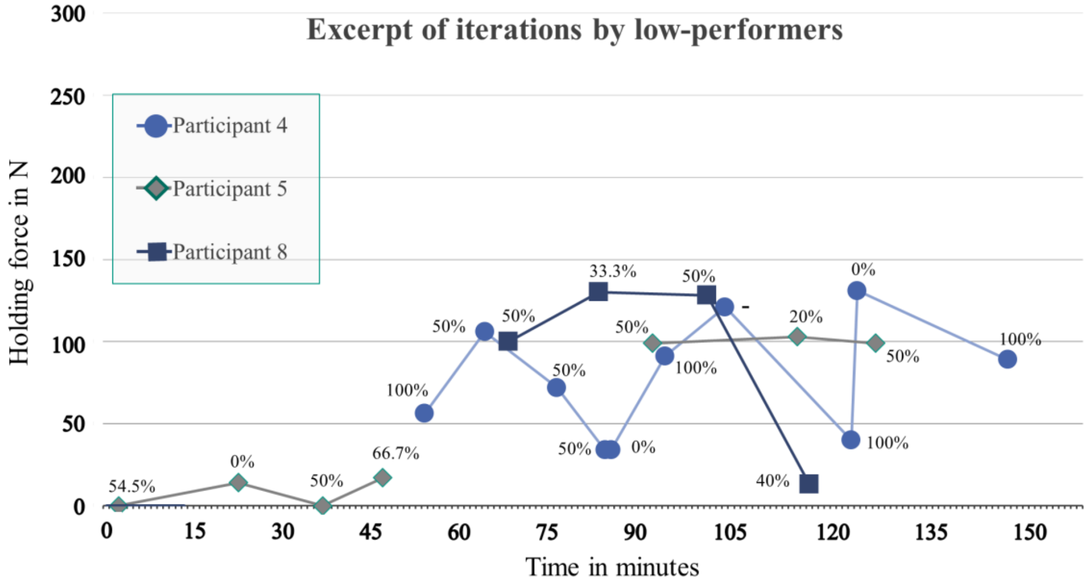

Figure 3. Holding force and percentage of function-relevant parameters in the changed parameters in iterations of the low-performers

The first pattern showed no increasing holding force over time. Participant 5 (gray rhombus) conducted four tests in the first 46 minutes, where the holding force reached a maximum of 17 N and was therefore a few N above the holding force of the starting configuration. The percentage of function-relevant parameters stayed relatively low during this pattern. Regarding participant 5, another period of constant holding force can be found. After 91 minutes, he reached a holding force of 100 N for the first time. This value could not be increased significantly over the next 40 minutes.

The next pattern shows oscillating holding forces over many iterations. Participant 4 (blue circles) relatively early (60 minutes) reached about 100 N, however, was unable to increase significantly even though he conducted many iterations afterwards. Here, a higher percentage of function-relevant parameters is shown, however even when 100% relevant parameters were changed, the force dropped due to change in an unsuitable direction.

The last pattern shows a sharp decrease in holding force of participant 8 (blue squares), that did not occur due to a destroyed snap-fit hook. The percentage of relevant parameters was never higher than 50%. Table 1 shows a median of 4 parameter changes per test for this participant.

In the following section, the high-performers (Figure 4) are investigated.

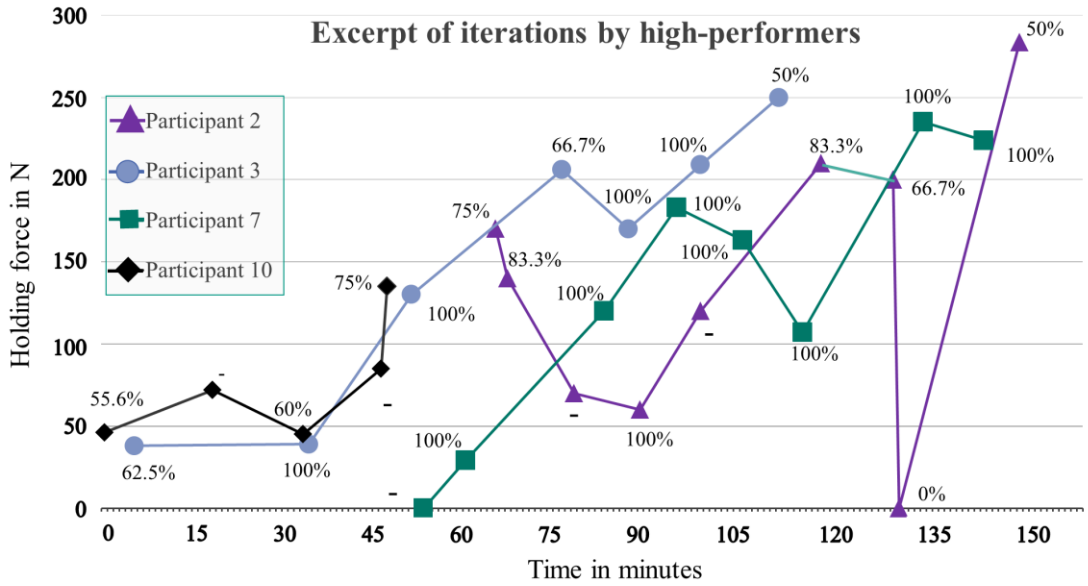

Figure 4. Holding force trend - percentage of function-relevant parameters in the changed parameters in iterations of the high-performers

https://doi.org/10.1017/pds.2023.221 Published online by Cambridge University Press

Participant 10 (black rhombus), 7 (green square) and 2 (lilac triangle) show repeated tests of the same design. No parameter was changed between the tests (marked by "-" instead of a % number). This is partially followed by a sharp increase in holding force.
The next pattern shows participant 3 (blue circles) who has a high rate of function-relevant parameters from the beginning. The curve of the measured holding force rises sharply after the second test. Only after about 90 minutes does the holding force decrease in one test and then increase to the maximum holding force of 250 N. Participant 3 changed a median of 2 parameters per iteration.
Participant 7 shows another pattern. He changed only relevant parameters in the iterations from the second test on. After about 50 minutes, the snap-fit connection broke once, in a test where no parameters were changed. From this point on, the holding force increased sharply. On median, participant 7 changed only one parameter per iteration.
Participant 2 (shown with purple triangles) also has a high rate of relevant parameters early on in the changed parameter between tests. A striking feature of Participant 2's curve is the continuous decline in holding force from a value above 150 N to nearly 50 N from 60 to 90 minutes. During this time period, Participant 2 started to repeat his tests without any changes in the parameters. Participant 2 and 7 destroyed their snap-fit connections, but were able to eliminate the problem.

4 DISCUSSION

Based on the results, the research question *What are successful approaches for testing activities to gain specific design knowledge?* can be answered as follows: During the investigation of the data from 10 participants, indications for successful and also less successful approaches to testing activities to gain specific design knowledge could be identified through categorization into high- and low-performers and analysis of their design decisions during the conducted iterations. We observed the following patterns:
Non-successful patterns: Stagnation of holding force, oscillating holding force or sharp decrease in holding force without destruction of the snap-fit hooks appeared primarily in low-performers. They are combined with low percentages of function-relevant parameters and high amount of changed parameters per iteration. These patterns are mostly descriptive, as they (except the high amount of parameters) cannot be influenced by supporting methods.
Successful patterns: A low amount of changed parameters in the iterations with a high percentage of function-relevant parameters combined with repeated tests without design changes and overstepping the structure's limit.
A non-influencing pattern was the amount of iterations, which seemingly contradicts results of Batliner et al. (2022), who indicate that more testing activities correlate with success in design. However, this might also result from the different study setups, as participants in the study of Batliner et al. (2022) were able to do many more product development activities in their semester project compared to the focus on testing in this study.
Based on the patterns for testing activities to gain specific design knowledge, the following hypotheses for successful approaches can be formulated:
Stagnation of holding force: Patterns, where the holding force does not increase over many iterations either through oscillation or stagnation indicate a low-performer. This leads to the hypothesis H1: *When the holding force shows a high frequency zigzag pattern or stagnates at the same level for several iterations, this indicates a low-performer.* The metric to identify and investigate this hypothesis could be e.g. that a moving trendline over three iterations should not be horizontal. This metric might be valuable, as it differentiates similar approaches between high- and low-performers. Participant 4 (low-performer) and 2 (high-performer) show partially a similar approach. The holding force value of participant 4 does not increase even though he was able to identify relevant parameters, but changed them in the "wrong" direction, as he seemed unable to clarify influence directions and interrelations of function-relevant parameters.
Amount of changed parameters per test: While wild guessing and fast testing with changes of many parameters at once seems to help at the beginning, it does not lead to success in the long term. Overdetermined test plans with more changed parameters per iteration do not lead to an improvement in holding force. This leads to the hypothesis H2: *A high amount of simultaneously changed parameters indicates a low-performer.* A metric might be the ratio of changed design parameters to

https://doi.org/10.1017/pds.2023.221 Published online by Cambridge University Press

conducted test runs. This can be explained through superposition of influences hindering distinctions of function-relevant and -irrelevant parameters.

Percentage of function-relevant parameters: Testing with a high proportion of relevant parameters and a low number of changed parameters leads to a sharp increase in the holding force achieved. That can be seen in Figure 4 in the graphs of participants 3 and 7. This is an expected observation, as it is a known issue for test engineers in design of experiments and many solutions are given in text books like Siebertz et al. (2017). However, in finding a solution to a design problem, these methods are often applied very late with aim of final optimization. In testing activities for gain of specific design knowledge, they are often left out, as design engineers are not necessary specialists in test planning. The hypothesis H3 is therefore as follows: *When the proportion of relevant parameters in the changed parameters is higher than the average, this indicates a high-performer.* A metric to test this hypothesis can be the moving average of the percentage of relevant parameters across three iterations.

Overstep the limit: A difference between High- and Low-performers emerges in tests with sharp drops in holding forces. The low performers' drop mostly does not correlate with a destroyed snap-fit connection, while high-performers break their systems more often. Participant 8 generates a sharp drop, when five parameters were changed and by changing one non-relevant parameter included to compensate for manufacturing tolerances, the snap fit hook was moved out of its counterpart and the holding force dropped by about 90%. The high-performer participant 2 shows a similar curve, where almost at the end, a drop to 0N occurs. Analysis shows that he made design decisions reaching the boundaries of mountability and breaks the snap fit hook before testing. However, in the next iteration, he is able to optimize the mounting without weakening the holding force, generating the global maximal force of 284N. *H5: When system failures alternate with high performance, the potential of the design is reached, which indicates a high-performer.* A metric for this hypothesis is the difference between the iteration before and after a failure.

Consider reproducibility: Comparative testing of the same configuration appears more often in high-performers. The repeated testing of the same configuration is observed in varying intensity in the graphs of all high-performers. An interesting indicator was that participant 2 (high-performer) was requesting new counterparts of the snap-fit connection after an iteration without changes in the design. This explains the strong increase in the holding force without changing the configuration between minutes 90 and 100 in Figure 4. The hypothesis H6 is therefore: *Running multiple tests with the same configuration without modification indicates a high-performer.* A metric is the holding force of the iteration before the comparison test divided by the holding force after the comparison test.

Limitations of this study are relatively low number of participants in the study, which limit the amount of identifyable patterns. Patterns might be misinterpreted or did not occur at all. This hampers derivation of hypotheses. Larger studies are needed, where the six hypotheses are investigated through their metrics and new patterns might emerge. This study also focusses on a very specific case of mechanical optimisation, where few components interact with each other. Therefore no generalisations can be to testing of systems with many interacting components or mechatronic components.

5 CONCLUSION AND OUTLOOK

Testing activities to gain specific design knowledge have been investigated in a laboratory environment with 10 participants. Hypotheses and corresponding metrics were derived as indications for successful and non-successful approaches to these testing activities. Four successful approaches like overstepping the structures limit, where participants accept the destruction of their system to identify boundaries, were identified. Also, three non-successful approaches like the amount of changed parameters in later tests were determined. As successful approaches are difficult to identify from the acquired raw data, especially in larger studies, key characteristics can be developed from our metrics in order to investigate successful approaches in more detail. With existing key characteristics, tailored support in terms of methods can then be developed and validated.

REFERENCES

Barhoush, Y.A.M., Erichsen, J.F., Sjöman, H., Georgiev, G.V. and Steinert, M. (2019), "Capturing Prototype Progress in Digital Fabrication Education", *Proceedings of the Design Society: International Conference on Engineering Design*, Vol. 1 No. 1, pp. 469–478.

Batliner, M., Boës, S., Heck, J. and Meboldt, M. (2022), "Linking Testing Activities with Success in Agile Development of Physical Products", *Procedia CIRP*, Vol. 109, pp. 146–154.
Cash, Philip; Stanković, Tino; Štorga, Mario (2016): Experimental Design Research. Cham: Springer International Publishing.
Clarkson, P.J., Wynn, D.C. and Eckert, C.M. (2007), Modelling iteration in engineering design, *16th International Conference on Engineering Design*, Paris, France.
Dorst, K. and Cross, N. (2001), "Creativity in the design process: co-evolution of problem–solution", *Design Studies*, Vol. 22 No. 5, pp. 425–437.
Engel, A. (2010), *Verification, validation, and testing of engineered systems, Wiley series in systems engineering and management*, J. Wiley & Sons, Hoboken.
Erichsen, J.F., Wulvik, A., Steinert, M. and Welo, T. (2019), "Efforts on Capturing Prototyping and Design Activity in Engineering Design Research", *Procedia CIRP*, Vol. 84, pp. 566–571.
Gero, J.S. and Kannengiesser, U. (2004), "The situated function–behaviour–structure framework", *Design Studies*, Vol. 25 No. 4, pp. 373–391.
Grauberger, P., Voß, K. and Matthiesen, S. (2019), "Functional Analysis in Embodiment Design - An Investigation of Embodiment Function Relations in Testing Activities", *Proceedings of the Design Society: International Conference on Engineering Design*, Vol. 1 No. 1, pp. 1503–1512.
Hubka, V. and Eder, W.E. (1990), "Design Knowledge: Theory in Support of Practice", *Journal of Engineering Design*, Vol. 1 No. 1, pp. 97–108.
IEEE (1990), *IEEE Standard Glossary of Software Engineering Terminology* No. IEEE-Std-610.12-1990, Institute of Electrical and Electronic Engineers, New York.
Kroll, E. and Weisbrod, G. (2020), "Testing and evaluating the applicability and effectiveness of the new idea-configuration-evaluation (ICE) method of conceptual design", *Research in Engineering Design*, Vol. 31 No. 1, pp. 103–122.
Matthiesen, S. (2021), "Gestaltung – Prozess und Methoden", in Bender, B. and Gericke, K. (Eds.), *Pahl/Beitz Konstruktionslehre*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 397–465.
Meboldt, M., Matthiesen, S. and Lohmeyer, Q. (2013), The Dilemma of Managing Iteration in Time-to-market Development Processes, *Proceedings of the 2nd International Workshop on Modelling and Management Engineering Processes*, Cambridge, United Kingdom
Shabi, J., Reich, Y. and Diamant, R. (2017), "Planning the verification, validation, and testing process: a case study demonstrating a decision support model", *Journal of Engineering Design*, Vol. 28 No. 3, pp. 171–204.
Siebertz, K., van Bebber, D. and Hochkirchen, T. (2017), *Statistische Versuchsplanung*, Springer Berlin Heidelberg, Berlin, Heidelberg.
Tahera, K., Wynn, D.C., Earl, C. and Eckert, C.M. (2019), "Testing in the incremental design and development of complex products", *Research in Engineering Design*, Vol. 30 No. 2, pp. 291–316.