# Investigating Modeling in Early Robust Design – a Study with the Modeling Method of Embodiment Function Relation and Tolerance Model

Jiahang Li[1], Dennis Horber[2], Philip Renner[1], Jiajie Yan[1], Stefan Goetz[2], Partric Grauberger[1], Sandro Wartzack[2], Sven Matthiesen[1]

[1]Karlsruhe Institute of Technology, Institute of Product Engineering, Germany
[2]Friedrich-Alexander-Universität Erlangen-Nürnberg, Engineering Design, Germany

**Abstract:** Robust Design (RD) aims to design systems that are insensitive to variations. While early RD methods can reduce costly iterations in product development, their evaluation often relies on case studies. Empirical investigation into early RD methods by applicants is still lacking, which hinders further research on the success criteria of the methods. This paper investigates the applicability and efficacy of the modeling method of the EFRT model through a cross-over study, comparing participants' results on robustness evaluation tasks. The result shows that the method is applicable and can support early robust design. The results provide a basis for future research and development of RD methods.

*Keywords: Empirical Studies, Robust Design, Design Models, Design Methods*

## 1 Introduction

Today's market requires less product development time and then the first right principle by design, which is difficult due to various variations in development. Robust design (RD) aims to design systems that are insensitive to variation (Taguchi et al., 2005). Traditional contributions to RD have primarily concentrated on optimizing robustness through controlled experiments in later design stages (Phadke, 1989; Taguchi et al., 2005), neglecting the significant impact of early design decisions (Jugulum and Frey, 2007).

To support early RD, multiple methods were developed. Mathias et al. (2011) use a three-level robustness ratio to analyze the influence of disturbances based on physical effects. Göhler and Howard (2015) considered contradictory goals and proposed a contradiction indicator (CI) for robustness evaluation. These approaches help by identifying and evaluating the problems without supporting the concrete design tasks. For engineering design tasks, common practices involve applying design principles like load path reduction and overdetermination avoidance (Andersson, 1997; Ebro et al., 2012; Ebro and Howard, 2016). Due to the lack of consideration of detailed information in individual product concepts, their application in engineering is challenging. The robustness of a product concept can be evaluated with the Embodiment Function Relation and Tolerance (EFRT) model by Horber et al. (2022), applying appropriate evaluation criteria, suggested by Li et al. (2023). For building up the EFRT model, a modeling method was developed (Li et al., 2024).

Eifler and Schleich (2021) conducted a review of RD methods, highlighting a notable gap in the investigation of their practical use. Typically, RD methods are mainly evaluated through individual case studies. For instance, Mathias et al. (2011) demonstrated robustness ratios using a wristwatch, while Göhler and Howard (2015) applied an injection device for the contradiction indicator (CI). Similarly, the modeling method of the EFRT model, as proposed by Li et al. (Li et al., 2024), was initially evaluated through case studies involving an angle grinder and a bicycle clipless pedal. From a broader perspective in design research, design methods in general lack sufficient empirical investigation (Barth et al., 2011; Eisenmann et al., 2021). This emphasizes the need for a more comprehensive understanding of the use of RD methods through empirical studies.

To investigate a design method, different approaches exist. Recognizing the need for comprehensive validation, existing approaches such as descriptive studies within the Design Research Methodology (Blessing and Chakrabarti, 2009) and the validation square (Pedersen et al., 2000) emphasize early investigations in a controlled environment, with a specific focus on efficacy. Efficacy is defined as the direct impact of implementing a design method on a designer's behavior (Daalhuizen and Cash, 2021). Besides confirming the efficacy of a design method, it is imperative to assess its applicability (Blessing and Chakrabarti, 2009). Eisenmann et al. (2021) suggest that applicability and efficacy should be investigated in the early stages of design method validation, before further development and validation of the method in subsequent steps.

Several empirical studies have been conducted on design methods. These studies can be adapted for the investigation of the RD methods. For example, Kroll and Weisbrod (2020) use a design experiment to evaluate the new idea-configuration-evaluation method by examining the design performance of participants. In addition, Grauberger et al. (2022) introduce a cross-over study design to evaluate the efficacy of the Contact and Channel Approach (C&C$^2$-A). In contrast to classical experiments with control and test groups, which generate one data point per participant, the cross-over study design allows for the double use of participants. This optimizes the use of resources and facilitates comprehensive comparisons within and between groups, increasing the depth of data analysis (Grauberger et al., 2022).

Motivated by the lack of investigation of RD methods, this paper aims to evaluate RD methods in an empirical study. We chose the EFRT model, because it enables the robustness evaluation of the product concepts and can support the design engineers in the early stages of product development. The problem is that the efficacy and applicability of the modeling method with the EFRT model have only been evaluated in case studies, there is still a lack of empirical investigation of the method by novice applicants, which hinders further research on its success criteria. To investigate the applicability and efficacy of the modeling method with the EFRT model, a cross-over study design with robustness evaluation tasks is conducted in this contribution, which is described in section 2. In the study, the applicability will be investigated using a questionnaire. The research will answer the research question below. In addition, the efficacy will be investigated with measurable data and criteria. The research hypothesis below should be tested by the results of the robustness evaluation tasks.

*Research question for applicability: how do participants assess the applicability of the modeling method with the EFRT model?*

*Research hypothesis for efficacy: the efficacy is shown by the better results of the robustness evaluation tasks in the study when the modeling method of the EFRT model is applied compared to intuitive approaches.*

## 2 Materials and methods

To investigate the applicability and efficacy of the modeling method with the EFRT model, a cross-over study was employed in this paper, building upon the study design proposed by Grauberger et al. (2022), with necessary adjustments to tasks and content.

The study design, depicted in Figure 1, involves an online course utilizing the ILIAS platform of the university of the authors' institute. After a brief introduction to robust design, participants are randomly assigned to two equally sized groups, A and B, each serving as a control group before and a test group after training in the EFRT model. The training unit is video-based and includes a questionnaire to assess the participant's understanding of the modeling method. At the end of the online study, a questionnaire on the modeling method is presented, consisting of two questions asking for the participant's subjective evaluation of the applicability of the modeling method by the participants. Efficacy is assessed by comparing participants' archived scores on the control and test group tasks. Applicability is assessed through questionnaires on the training unit and modeling method. The study is conducted in German, but for the clarity of this paper, all information is translated into English. The study questions are arranged chronologically as follows, each question has a time limit of 10 minutes, and there is a total time limit of 80 minutes for the study, with participants receiving no feedback on their answers throughout the process.

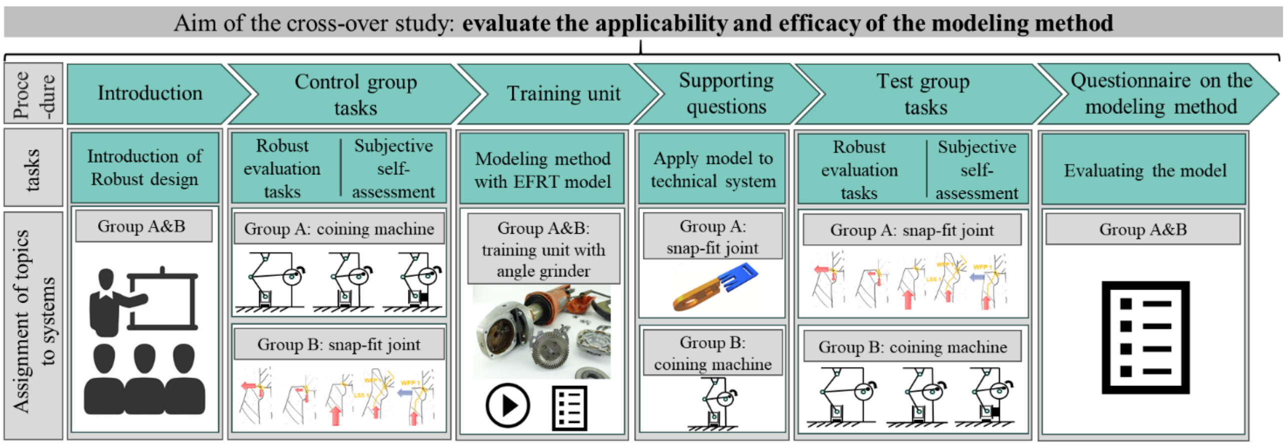



Figure 1. Overview of the cross-over study design

Two technical systems, a snap-fit joint and a coining machine, are utilized for the control group and test group tasks. Participants initially handle control group tasks focusing on the system behavior of one technical system, with groups A and B managing different control group tasks separately. The training unit is based on pre-recorded videos to minimize influences from study conductors and includes an application example on an angle grinder. After the training unit, participants proceed to test group tasks by switching systems between groups A and B. The type of tasks for the two technical systems remains the same, ensuring the consistent application of the modeling method. In this study, group A is assigned the coining machine task in the control group and then the snap-fit joint, while Group B follows the reverse order. The training unit as well as the control and test group tasks are described in detail in sections 2.1 and 2.2.

### 2.1 Training unit in the modeling method with the EFRT model

The aim of the training unit is for the participants to understand the modeling method and to apply it to the test group tasks. For the validation of a new method, a clear and detailed description of the method and its intended use is essential (Gericke et al., 2017). Therefore, the modeling method is short introduced here. The modeling method with the EFRT model outlines a systematic approach for modeling product concepts to assess their robustness. In Figure 2, an overview of the modeling method within a stage-gate process is provided. It comprises five stages and gates.

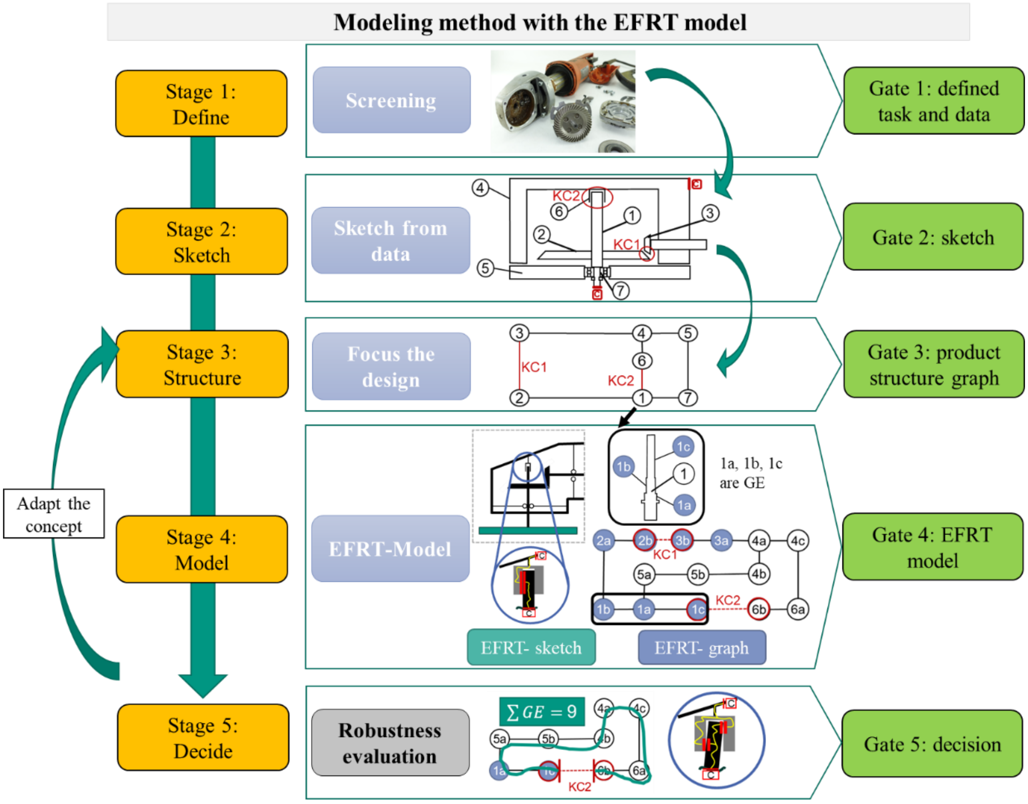



Figure 2. Modeling method with the Embodiment Function Relation and Tolerance (EFRT) model according to Li et al. (2024). KC: Key Characteristic. GE: Geometry Element.

The steps of the modeling method are explained through five training videos and a comprehensive handout provided to participants in both digital and printed format. In the training unit, a real angle grinder system is first introduced to the participants. In the first stage, the task is defined as reducing the rejection rate in the assembly. In stage 2, the angle grinder is simplified in a sketch with the components relevant to the task. Then, in stage 3, the relationships of these components are represented in the product structure graph. In stage 4, the EFRT model of the angle grinder is built, including an EFRT graph to describe the structure of the product concepts and an EFRT sketch to visualize the concept. In the EFRT graph, the product concept is deconstructed into small geometric elements (GE) and their geometric relationships. The EFRT sketch supplements the product sketch with key elements of the C&C$^2$-A, enabling the analysis of system behavior in different system states. Key characteristics (KCs) are used in the robustness evaluation to analyze the target function of the product. The influence of different product concepts on the KCs can be analyzed using the EFRT graph and the EFRT sketch. The built EFRT model can be used for robustness evaluation with appropriate criteria in stage 5.

Each stage of the modeling method in the training unit is accompanied by a training video that explains the steps for building the model. After each training video, there is a questionnaire to assess the participant's understanding of the modeling method. The questionnaire asks how understandable the stages in the video were. Four possible answers are available (very good, rather good, rather bad, very bad). The evaluation principle allocates a maximum of four points for each question if the answer is very good, and at least one point if the answer is very unsatisfactory.

After the training videos, supporting questions for each stage related to the second technical system are provided for robustness assessment preparation using the modeling method with the EFRT model. The purpose of these questions is to support the participants in building the EFRT model for the test group tasks with the modeling method. Participants don't receive any feedback about these questions.

## 2.2 Control and test group tasks

As efficacy is assessed by comparing participants' achieved scores on the control and test group tasks, it is necessary to introduce the basic structure of the control group and test group tasks. To investigate efficacy, each task is divided into two steps, each accompanied by a question. The first step focuses on the correctness of the robustness evaluation, providing an objectively assessable metric for the evaluator. Then, in the second step, participants are asked about their confidence in their answer selection, collecting the subjective self-assessment of the study participants, see Figure 3.

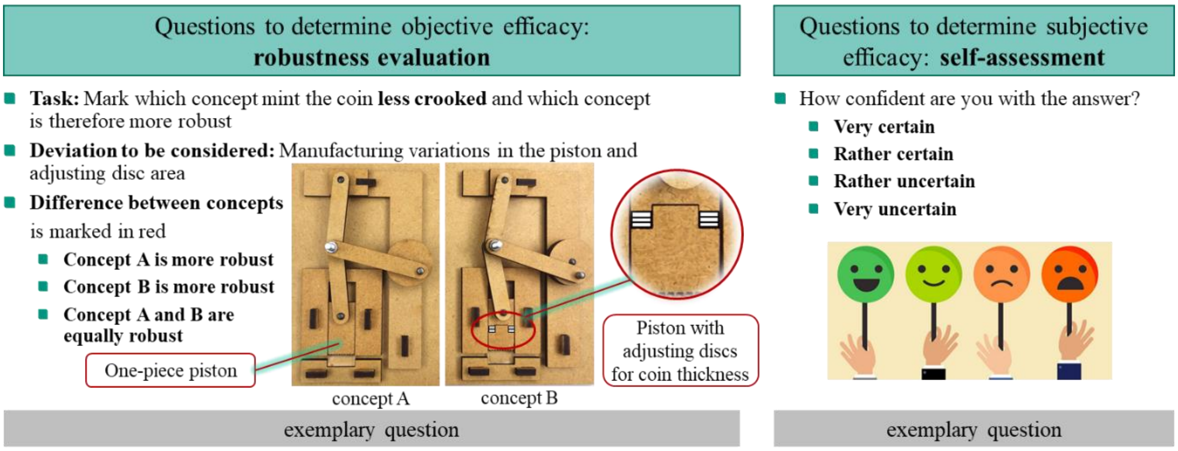



Figure 3. Task structure of the control group and test group tasks

In the first step, participants are presented with a task comparing two concepts, A and B, with a clear development goal provided (e.g., coining machine "crooked minting of the coin has to be improved"). An exemplary question in this step is depicted in Figure 3 left, participants must decide which concept is more robust based on various design parameters, represented by photos or principle sketches. The differences between the two concepts are highlighted, and the deviations to be considered are provided. In the example above, the concept of minting the coin less crooked under manufacturing variation is considered more robust. Applying the evaluation criteria "load path" proposed by Li et al. (2023) leads to the conclusion that concept A is more robust, since there is less manufacturing variation of the components in the load path that contribute to the minting angle. Participants select one of three answers, as depicted in Figure 3. The evaluation principle awards a maximum of one point if the question is answered correctly and zero points if incorrect. The summed scores reflect the correct evaluations given by the participants and range from 0 to 7 points, as there are a total of 7 robustness evaluation questions each for control and test group tasks.

In the second step, participants engage in a self-assessment regarding their confidence in the chosen answer on a four-point scale (very certain, rather certain, rather uncertain, very uncertain). The evaluation principle allocates a maximum of four points for each question if the participant is very certain and at least one point if very uncertain. This structured approach ensures a comprehensive evaluation of both objective and subjective efficacy, contributing to a thorough understanding of the participant's engagement with the tasks in the control and test groups.

## 2.3 Participants

The participants in this study are students enrolled in the fifth semester of the mechatronic bachelor course at the authors' institute. As part of their coursework, these students are engaged in an engineering design project within the program. They possess a foundational understanding of the product development process. In addition, they have received lectures on modeling that have provided them with a basic understanding of how to build models for development tasks, e.g., using C&C²-A. However, these students have not been exposed to lectures on modeling with the EFRT model before participating in this study. The students were informed that the study would not influence their performance in the course and the participation was voluntary. Additionally, the test is pseudonymized. No personal data, such as the gender or age of the participants is collected during the study.

In this paper, we consider students as participants as they have a homogeneous educational background. According to Carver et al. (2010), using empirical studies with engineering students helps researchers gain insight into new or existing methods. One advantage is that students have the academic background necessary to work with design methods that can meet the needs of the researcher (Üreten et al., 2021). It should be noted that students may not fully represent the professionals in the field, as their skill levels may differ from those of experienced practitioners (Feldt et al., 2018). Since the students in the study have taken engineering courses and have worked on an engineering design project in the course, it can be considered sufficient to include them as participants.

### 2.4 Data analysis

The procedure for data analysis is illustrated in Figure 4. Initially, the data is filtered using two criteria. Criterion 1 addresses the completion of the study and any non-completed samples were removed. Criterion 2 addresses successful participation in the method training. After each training video, participants were asked a question about the video content. The correct answer was shown in the video. This helps ensure that people who didn't answer correctly weren't paying close attention to the videos. Therefore, an incorrect answer will result in the removal of the samples.

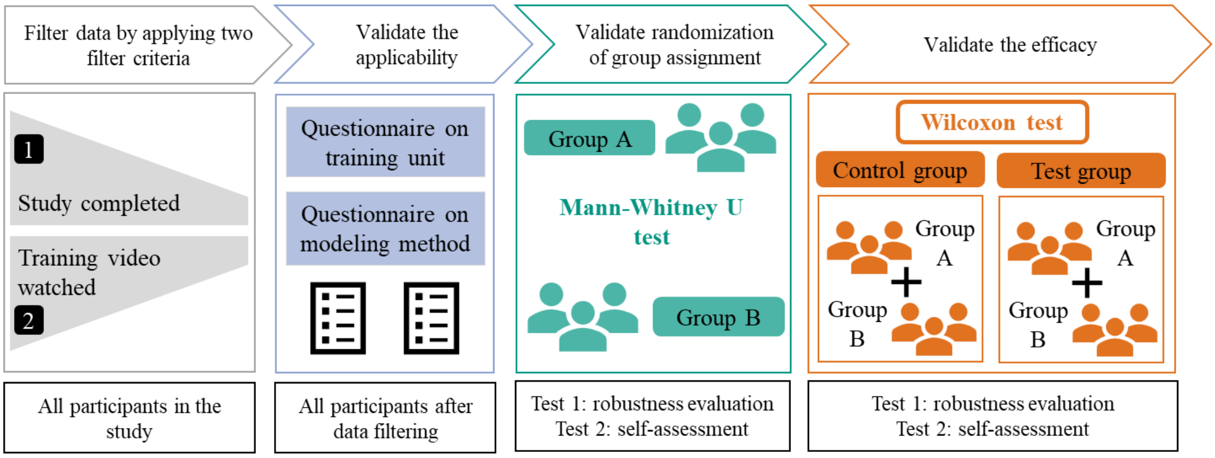



Figure 4. Data analysis process to validate the applicability and efficacy of the modeling method with the EFRT model

The primary focus of the initial data quality analysis is to assess whether the modeling method can be applied by the participants. The applicability is assessed with the questionnaires in the training unit and in the modeling method. Since the questions are independent of the group division, the results of the two groups are analyzed together and evaluated using boxplots.

For the efficacy of the modeling method, the scores of the control group tasks and test group tasks are calculated and prepared for statistical tests. The procedure for the evaluation of robustness assessment and self-assessment remains identical. Mann-Whitney U test and Wilcoxon test are selected for the statistical data analysis, with the statistical tests conducted using SPSS software. The Mann-Whitney U test compares the summarized scores of the control group tasks and test group tasks between groups A and B. This test assesses potential effects within participant groups and validates the randomization of sorting participants into Groups A and B. If randomization is confirmed, both groups can be treated as one control group before the method training and one test group after the training. The Wilcoxon test, designed for dependent samples, is utilized to analyze differences between the control and test groups, considering that the data were collected from the same participants. The scores achieved in control group tasks and test group tasks are then compared with the Wilcoxon test to evaluate the efficacy of the modeling method. In both tests, there will be a significant difference if the p-value is less than 0.05.

## 3 Result

The study was conducted in a lecture hall and Figure 5 provides an overview of the study implementation.

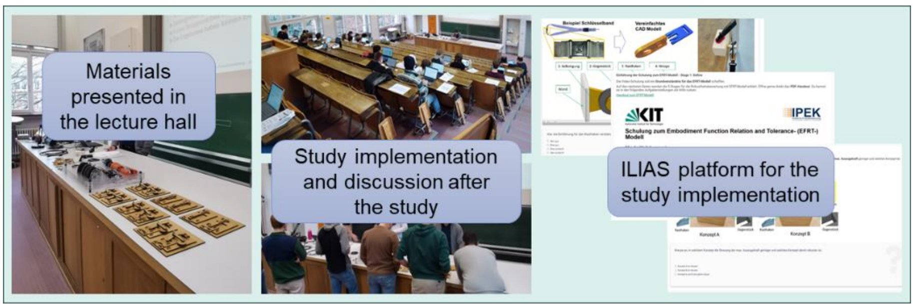



Figure 5. Overview of the study implementation and participants

A total of 20 participants took part in the test. Out of the 20 participants, the data of two participants were filtered according to criterion 1, and the data of one participant were filtered according to criterion 2 as described in section 2.4. Consequently, a total of 17 participants (N = 17) were included in the subsequent analyses, while 9 participants belong to group A and 8 participants belong to group B. Each participant contributed two data points (control and test group), which were included in the subsequent analyses for the efficacy of the modeling method.

### 3.1 Applicability of the modeling method with the EFRT model

The applicability of the modeling method is evaluated through two questionnaires, the results of which are described below. All results are based on the subjective assessment of the study participants.

#### 3.1.1 Evaluation of the questionnaire on the training unit

The evaluation of the five questions about the training unit focused on the study participants' assessments of their understanding of the explanation provided in the training video for each of the five stages of the modeling method with the EFRT model. The result is depicted in Figure 6, the boxes represent the interquartile range from the lower to the upper quartile. It includes the median scores with the black lines and the mean scores with the cross marks. Overall, participants rated each of the five stages as being presented in an understandable way, with a score above 3 out of 4. The lowest rating was given to Stage 1, where the explanation in the training video was perceived as slightly less clear than the other four stages. Despite this, participants generally found the content of the stages in the modeling method to be understandable.

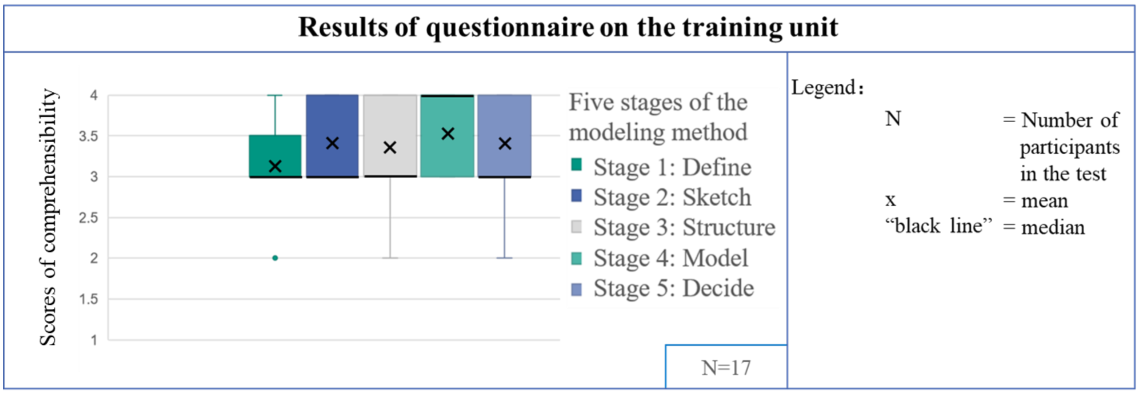



Figure 6. Results of the questionnaire on the training unit show the participants' positive assessments of their understanding of the five stages of the modeling method

#### 3.1.2 Evaluation of the questionnaire on the modeling method

The questionnaire on the modeling method includes two questions regarding how the study participants assess the applicability of the modeling method with the EFRT model and their overall assessment of the modeling method. The response to the questionnaire is given as a percentage, with 0 percent as the worst and 100 percent as the best rating. The questions and results are shown in Figure 7.

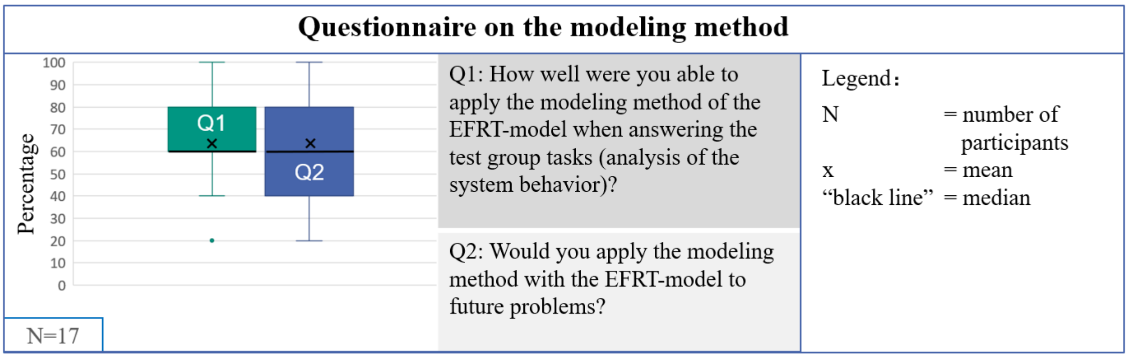



Figure 7. Results of the questionnaire on the modeling method show the participants' positive assessments

The evaluation of the first question Q1 suggests that the study participants found the model easy to apply to solve the test group's tasks (M = 63.53%, Mdn = 60%, first quartile = 60%, third quartile = 80%, SD = 17.1%). This means that the majority of participants believe that they could apply the modeling method to the test group tasks. Responses to the second question Q2, the overall evaluation of the modeling approach, were positive (M = 63.53%, Mdn = 60%, first quartile = 40%, third quartile = 80%, SD = 24.0%). This means that the majority of the participants consider the modeling method to be helpful and that they would consider using the modeling method to solve future engineering problems.

### 3.2 Efficacy of the modeling method with the EFRT model

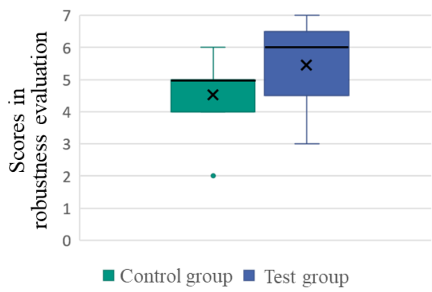
The efficacy of the modeling method is evaluated through control and test group tasks. As described in section 2.2, each task has two steps: the first step addresses the objective efficacy, and the second focuses on the subjective efficacy.

#### 3.2.1 Objective efficacy of the participants using the modeling method of the EFRT model

The results for the first step of the control and test group tasks are described below. The summed scores between Group A and Group B are first compared using the Mann-Whitney U test. The exact Mann-Whitney U test results in U = 19, p = 0.114. This means that group A and group B don't have a significant difference in robustness evaluation tasks.

The evaluation of objective efficacy is depicted in Figure 8. Since the group comparison shows no significant effect, the data from both groups can be treated as one data set for the Wilcoxon test comparing the data between the control group and the test group. The Wilcoxon test results in a p-value of 0.03, indicating a significant difference. The effect r is calculated as 0.528, which is a strong effect according to Cohen (2013). As shown in Figure 8, the test group achieved a significantly higher score than the control group, indicating a positive objective efficacy in the robustness evaluation.

**Objective efficacy of the modeling method**




| Test statistics - Wilcoxon test | |
|---|---|
| N | 17 |
| Z | 2.176 |
| p | 0.030 |
| r | 0.528 |

Legend:

| | |
|---|---|
| N | = number of participants in the test |
| Z | = standardized random variable |
| P | = value of significance |
| r | = effect according to Cohen (2013) |
| x | = mean |
| “black line” | = median |

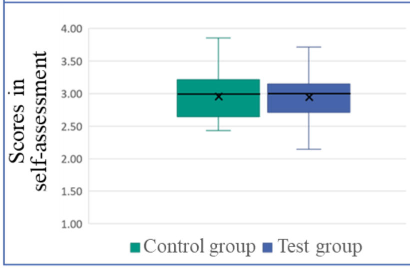
Figure 8. Results in the control and test group tasks for objective efficacy based on robustness evaluation

#### 3.2.2 Subjective efficacy of the participants using the modeling method of the EFRT model

The results for the second step of the control and test group tasks are described below. Using the Mann-Whitney U test, Group A showed no significant difference in scores compared to Group B, as indicated by an exact Mann-Whitney U test result of U = 20, p = 0.121. This means that the randomly assigned groups A and B don't have a significant difference in their confidence levels in answer selection.

The evaluation of subjective efficacy is depicted in Figure 9. To further compare the control and test groups, Wilcoxon tests were performed. However, there was no significant difference between the control and the test group (N = 17, p = 0.975). The results indicated that the subjective efficacy showed no significant differences when comparing their confidence levels in answer selection.

**Subjective efficacy of the modeling method**




| Test statistics - Wilcoxon test | |
|---|---|
| N | 17 |
| Z | -0.032 |
| p | 0.975 |

Legend:

| | |
|---|---|
| N | = Number of participants in the test |
| Z | = Standardized random variable |
| P | = Value of significance |
| x | = mean |
| “black line” | = median |

Figure 9. Results in the control and test group tasks for subjective efficacy based on participants’ self-assessment

## 4 Discussion

Based on the results of the questionnaires in sections 3.1.1 and 3.1.2, the research question "*How do participants assess the applicability of the modeling method with the EFRT model?*" can be answered as follows.

The applicability assessed through the questionnaires on the training unit (section 3.1.1) and modeling method (section 3.1.2), shows a sound applicability of the modeling method. The evaluation of the questions about the training unit in section 3.1.1 reveals that the participants understood the method, as evident from their responses regarding the clarity of the stages in the modeling method. As depicted in Figure 6, all stages are rated with a score above 3 out of 4, this indicates that the training was overall successful, and the method is perceived as understandable. Figure 6 also shows, which stage of the method is providing the participants with a better understanding of the training unit. This indicates the direction of potential adaptation in the explanation of each stage of the modeling method for further study design.

The participants' subjective assessment of applicability was also reflected in the results of the questionnaire on the modeling method in section 3.1.2. As shown in Figure 7, the response to the first question (Q1) indicates that study participants have a positive view of the applicability of the method for solving the test group tasks ($M = 63.53\%$, $SD = 17.1\%$). The result of the second question (Q2) showed that the participants also expressed a willingness to use the EFRT model in the future, which indicates the perceived applicability of the modeling method ($M = 63.53\%$, $SD = 24.0\%$).

With the questionnaires designed in the study, it is possible to get quick feedback and an initial picture of the applicability of an RD method, which is important for its further validation (Blessing and Chakrabarti, 2009). In the questionnaires, the subjective assessment of the participants was quantified in scales, which facilitates the evaluation of the results, but also influences the validity of the results (Ouwehand et al., 2021). For more insight into how to improve the applicability, an interview with the participants, as conducted by Kroll and Weisbrod (2020), may be helpful.

Based on the results of the control and test group tasks in section 3.2.1 and 3.2.2, the research hypothesis "*The efficacy is shown by the better results of the robustness evaluation tasks when the modeling method of the EFRT model is applied compared to intuitive approaches*" can be partially confirmed and is discussed below.

We found that the results from the test group tasks show significant improvement compared to the control group tasks in robustness evaluation ($p = 0.03$), see Figure 8. In the cross-over study design, this effect is considered to be caused by the application of the modeling method, as the participants have successfully completed the training unit in the modeling method. Following the five stages of the modeling method, the participants are supported in systematically analyzing the task and problem, focusing on the deviations, and evaluating their impact on the target function of the system, which is the basis for the robustness evaluation. Combined with the participants' subjective assessment that they can apply the modeling method to the robustness evaluation tasks (Q1 in Figure 7), the results of the robustness evaluation tasks show that the objective efficacy of the modeling method is positive.

The subjective efficacy showed no significant differences when comparing their confidence levels in answer selection ($p = 0.975$), see Figure 9. The results present a notable discrepancy when compared to objective efficacy. This discrepancy contrasts with the positive assessment of the modeling method expressed in the questionnaire. This raises the following question: Why do the scores for the self-assessment of the study participants after the training unit not increase, even though they express positive views on the applicability of the method in the questionnaire on the modeling method?

One possible reason for this difference is the insufficient time to consolidate the content of the method. The time frame may have been too short for participants to internalize the content of the method and build confidence in its application. This effect also addresses the identified challenges in investigating the long-term effects of the method with the cross-over design (Grauberger et al., 2022). This could be better investigated by conducting a long-term study, allowing participants more time to learn and become familiar with the EFRT model over a longer period. Another reason may be the lack of feedback during training. Participants did not receive feedback during training on the modeling tasks, which led to uncertainty in their responses to the test group tasks. Providing feedback after the modeling tasks could increase participants' confidence in their understanding and application of the modeling method. Furthermore, the method may require participants to consider additional issues, e.g., identification of new problems, which increases complexity and potential uncertainty.

Based on the results, a further step is to examine and revise the study design before modifying the method itself. This is due to the fact that the self-assessment scores should increase in the test group tasks. This can be achieved by revising the training unit. The focus should be on increasing participants' confidence in applying the EFRT model. Concretely, this could mean that study participants receive direct feedback on questions related to the training unit. This allows them to compare their results with the correct answer and thus deepen their learning. Another result of the study is that the study participants were almost below the planned processing time of 80 minutes. Instead of shortening the overall processing time, an extension of the training unit could be more advisable. Through further modeling tasks that not only test understanding, the study participants can independently apply what they have learned. The hypothesis that a modified study design in the form of an adaptive training unit helps to strengthen the self-assessment of study participants when using the EFRT model needs to be tested in a further study. Based on the results obtained with the newly adapted study, the modeling method with the EFRT model may be improved in a further step. In addition to further investigation of the modeling method with EFRT model, the insights gained from this study also provide references for the research of design methods in other disciplines.

Due to the lack of investigation of RD methods identified by Eifler and Schleich (2021), the applicability and efficacy of the modeling method of the EFRT model was now initially evaluated in an empirical study, which is a foundation before further research or application in practice (Eisenmann et al., 2021). Compared to the evaluation with the case studies, e.g., by Mathias et al. (2011) and Göhler and Howard (2015), RD methods can now also be evaluated with novice applicants. In addition, this study design allows for the comparison of different robust design methods. This may provide more meaningful insights than a comparison with the intuitive approaches of the participants in the control group tasks, given the uncontrollable variety of approaches and the resulting uncertainty about the causes of the individual scores.

The paper acknowledges several limitations. First, the exclusive focus on two specific technical systems limits the generalizability of the results. It is recommended that the applicability and efficacy of the modeling method be evaluated on a broader range of industry-related systems. The limited number of participants in the study, coupled with the lack of a normal distribution and few data points, reduces the statistical significance of the observed differences after training. In addition, the study design lacks the ability to analyze the individual weaknesses of each stage of the modeling method, suggesting the potential for more detailed examination in future research.

## 5 Conclusion

Motivated by the lack of investigation of RD methods with empirical studies, this paper explores the efficacy and applicability of the modeling method with the EFRT model for early robustness evaluation through a cross-over study. The modeling method is introduced to the participants with the video-based training unit. Efficacy is assessed by comparing participants' archived scores on tasks without and with successful participation in training in the method. The results of the study showed that the scores on the robustness evaluation tasks increased significantly with the method training, which indicates a positive objective efficacy of the modeling method. The subjective efficacy showed no significant differences when comparing their confidence levels in answer selection. The applicability assessed through the questionnaire on the training unit and modeling method shows a sound applicability of the modeling method. This means that the modeling method of the EFRT model is applicable and can support early robust design, even if the applicants may not perceive immediate success. Based on the results of the study, improvements to the study are suggested, such as integrating more feedback into the method training to increase participants' confidence level while using the method, encouraging further research on RD methods.

## 6 Acknowledgements

We would like to thank all participants for taking part in the studies. The research documented in this manuscript has been funded by the German Research Foundation (DFG) within the project "Holistic robustness evaluation in early design stages" (467789897). The support by the German Research Foundation (DFG) is gratefully acknowledged.

## 7 References

Andersson, P., 1997. On Robust Design in the Conceptual Design Phase: A Qualitative Approach. Journal of Engineering Design 8, pp. 75–89. DOI:10.1080/09544829708907953.

Barth, A., Caillaud, E. and Rose, B., 2011. How to validate research in engineering design, in: Proceedings of the 18th International Conference on Engineering Design. ICED'11, Lyngby/Copenhagen, Denmark.

Blessing, L.T.M., Chakrabarti, A., 2009. DRM, a design research methodology. Springer, Dordrecht, New York.

Carver, J.C., Jaccheri, L., Morasca, S. and Shull, F., 2010. A checklist for integrating student empirical studies with research and teaching goals. Empirical Software Engineering 15, pp. 35–59. DOI:10.1007/s10664-009-9109-9.

Cohen, J., 2013. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Taylor and Francis, Hoboken, 567 pp.

Daalhuizen, J. and Cash, P., 2021. Method content theory: Towards a new understanding of methods in design. Design Studies 75, p. 101018. DOI:10.1016/j.destud.2021.101018.

Ebro, M., Howard, T. and Rasmussen, J., 2012. The foundation for robust design: Enabling robustness through kinematic design and design clarity. Proceedings of International Design Conference, DESIGN,

Ebro, M. and Howard, T.J., 2016. Robust design principles for reducing variation in functional performance. Journal of Engineering Design 27, pp. 75–92. DOI:10.1080/09544828.2015.1103844.

Eifler, T. and Schleich, B., 2021. A robust design research landscape - review on the importance of design research for achieving product robustness. Proceedings of the Design Society 1, pp. 211–220. DOI:10.1017/pds.2021.22.

Eisenmann, M., Grauberger, P. and Matthiesen, S., 2021. Supporting Early Stages of Design Method Validation - An Approach to Assess Applicability. Proceedings of the Design Society 1, pp. 2821–2830. DOI:10.1017/pds.2021.543.

Feldt, R., Zimmermann, T., Bergersen, G.R., Falessi, D., Jedlitschka, A., Juristo, N., Münch, J., Oivo, M., Runeson, P., Shepperd, M., Sjøberg, D.I.K. and Turhan, B., 2018. Four commentaries on the use of students and professionals in empirical software engineering experiments. Empirical Software Engineering 23, pp. 3801–3820. DOI:10.1007/s10664-018-9655-0.

Gericke, K., Eckert, C. and Stacey, M., 2017. What do we need to say about a design method, in: Proceedings of the 21st International Conference on Engineering Design. ICED'17, Vancouver, Canada, pp. 101–110.

Göhler, S.M. and Howard, T.J., 2015. The Contradiction Index (CI): A New Metric Combining System Complexity and Robustness for Early Design Stages, in: 27th International Conference on Design Theory and Methodology, Boston, Massachusetts, USA. 2015. DOI:10.1115/DETC2015-47255.

Grauberger, P., Eisenmann, M., Windisch, E. and Matthiesen, S., 2022. Investigating the generation of specific design knowledge: experimental validation of a qualitative modelling method. Journal of Engineering Design 33, pp. 870–895. DOI:10.1080/09544828.2022.2151788.

Horber, D., Li, J., Grauberger, P., Schleich, B., Matthiesen, S. and Wartzack, S., 2022. A model-based approach for early robustness evaluation – Combination of Contact and Channel Approach with tolerance graphs in SysML, in: Proceedings of the 33rd Symposium Design for X. DFX 2022, Hamburg.

Jugulum, R. and Frey, D.D., 2007. Toward a taxonomy of concept designs for improved robustness. Journal of Engineering Design 18, pp. 139–156. DOI:10.1080/09544820600731496.

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 31, pp. 103–122. DOI:10.1007/s00163-019-00324-6.

Li, J., Horber, D., Grauberger, P., Goetz, S., Wartzack, S. and Matthiesen, S., 2024. Supporting early robust design for different levels of specific design knowledge – An adaptive method for modeling with the Embodiment Function Relation and Tolerancing model. KIT Scientific Working Papers 246, Karlsruhe. DOI:10.5445/IR/1000170495.

Li, J., Horber, D., Keller, C., Grauberger, P., Goetz, S., Wartzack, S. and Matthiesen, S., 2023. Utilizing the Embodiment Function Relation and Tolerance Model for Robust Concept Design. Proceedings of the Design Society 2023, 3, pp. 3771–3780. DOI:10.1017/pds.2023.378.

Mathias, J., Kloberdanz, H., Eifler, T., Engelhardt, R., Wiebel, M., Birkhofer, H. and Bohn, A., 2011. Selection of physical effects based on disturbances and robustness ratios in the early phases of robust design, in: Proceedings of the 18th International Conference on Engineering Design. ICED'11, Lyngby/Copenhagen, Denmark, pp. 324–335.

Ouwehand, K., van der Kroef, A., Wong, J. and Paas, F., 2021. Measuring Cognitive Load: Are There More Valid Alternatives to Likert Rating Scales? Frontiers in Education 6, 702616. DOI:10.3389/feduc.2021.702616.

Pedersen, K., Emblemsvag, J., Bailey, R., Allen, J.K. and Mistree, F., 2000. The "Validation Square" - Validating Design Methods. Proceedings of DETC `00, 2000 ASME Design Engineering Technical Conferences, Baltimore, Maryland, USA,

Phadke, M.S., 1989. Quality Engineering using Design of Experiments, in: Dehnad, K. (Ed.), Quality Control, Robust Design, and the Taguchi Method. Springer US, Boston, MA, pp. 31–50.

Taguchi, G., Chowdhury, S., Wu, Y., Taguchi, S., Yano, H., 2005. Taguchi's quality engineering handbook. John Wiley & Sons, Hoboken, N.J., Livonia, Mich., 1662 pp.

Üreten, S., Sankowski, O. and Krause, D., 2021. Student participants in empirical studies in engineering design - a collection of reflections to improve your study quality. Proceedings of the Design Society 1, pp. 2741–2750. DOI:10.1017/pds.2021.535.

**Contact: Jiahang Li,** Karlsruhe Institute of Technology (KIT), IPEK – Institute of Product Engineering, Kaiserstr. 10, 76131 Karlsruhe, Germany, Phone: +49 721 608-42735, Fax: +49 721 608-46051, jiahang.li@kit.edu