

# Assessing the User Experience: A Usability Report on Delta3 GmbH Software

Akshay Chikhalkar

*Department of Electrical Engineering and Computer Science*  
*Technische Hochschule Ostwestfalen-Lippe University of Applied Sciences and Arts*  
Lemgo, Germany  
akshay.chikhalkar@stud.th-owl.de  
Matrikel-Nr: 15489036

## Abstract

This report explores software usability engineering, a critical aspect of software development that ensures the creation of easy-to-use, intuitive, and efficient software applications. It emphasizes the significance of software usability and identifies the factors that influence it. The report features a case study of Delta3 GmbH, a German software development company specializing in 3D scanning and modeling software. The study, conducted in both English and German, involved expert evaluation and usability testing for three software products developed by Delta3 GmbH. Participants from different professions, age groups, and educational backgrounds who were unfamiliar with the software were recruited for the study. The report presents the study's findings, identifying the strengths and weaknesses of the products and analyzing the user experience questionnaire.

## Index Terms

User experience questionnaire (UEQ), usability testing, heuristic evaluation, app/player, editor, operations, user observations.

## CONTENTS

<b>I</b>	<b>Introduction</b>	<b>3</b>
<b>II</b>	<b>Expert Evaluation</b>	<b>3</b>
II-A	Methods . . . . .	3
II-B	Procedures . . . . .	4
II-C	Results . . . . .	4
II-C1	App/Player . . . . .	4
II-C2	Editor . . . . .	5
II-C3	Operations . . . . .	5
<b>III</b>	<b>Usability Tests</b>	<b>6</b>
III-A	Methods . . . . .	6
III-A1	User study . . . . .	6
III-A2	User experience questionnaire . . . . .	6
III-B	Procedures . . . . .	7
III-C	Results . . . . .	7
III-C1	User observations . . . . .	7
III-C2	Mean value per item . . . . .	8
III-C3	UEQ scales . . . . .	9
III-C4	Pragmatic & Hedonic Quality . . . . .	10
III-C5	Benchmark . . . . .	10
III-C6	KPI calculation . . . . .	11
III-C7	Sample size . . . . .	11
<b>IV</b>	<b>Discussion</b>	<b>12</b>
<b>V</b>	<b>Recommendations</b>	<b>13</b>
	<b>Appendix A: Tables</b>	<b>14</b>
	<b>Appendix B: Figures</b>	<b>15</b>
	<b>References</b>	<b>26</b>

## I. INTRODUCTION

Software usability engineering is a critical aspect of software development that focuses on designing and developing software applications that are easy to use, intuitive, and efficient [1]. The success of a software product heavily depends on its usability and how well it meets the needs of its intended users [2]. Usability is determined by several factors, including interface design, instructions and documentation, performance, and reliability. Usability testing and expert evaluation are common methods used to assess the usability of software products and make necessary improvements [3].

In this report, we will explore the importance of software usability engineering in more detail. We will discuss the various factors that influence software usability and the methods used to evaluate and improve it. Additionally, we will delve into the specific case of Delta3 GmbH, a software development company based in Germany that specializes in 3D scanning and modeling software solutions for various industries. We will examine their software products' usability.

Furthermore, we will discuss the study conducted on the three software products developed by Delta3 GmbH, which is divided into two parts: expert evaluation and usability testing. The expert evaluation was based on 10 heuristic principles, while the user study involved five participants from different professions. We will discuss the results of the study, including the strengths and weaknesses of the software products and the improvements that can be made to enhance their usability.

## II. EXPERT EVALUATION

Expert evaluation is a method of usability testing in which usability experts review a software product or system and identify potential usability issues. This method is often used to complement other forms of usability testing, such as user testing, and can be performed at any stage of the software development process [4].

During an expert evaluation, one or more usability experts examine the software product and assess its usability based on established usability principles and guidelines. The experts may use various techniques to evaluate the software, such as heuristic evaluation, cognitive walkthrough, or task analysis [4], in this study we used heuristic evaluation.

The goal of expert evaluation is to identify potential usability issues early in the development process, so that they can be addressed before the software is released to users. By using expert evaluation to identify and address usability issues, software developers can improve the overall usability of their products, which can lead to increased user satisfaction and improved user performance [4].

### A. Methods

Heuristic evaluation is a method used to systematically evaluate the usability of a user interface or product by a small group of evaluators [4]. These evaluators consist of individuals with relevant experience and expertise in usability, interaction design, and related fields. The evaluation process involves examining the interface against a set of predefined usability heuristics or guidelines that have been established through research and experience [4].

During the evaluation process, the evaluators identify any usability issues that violate the established heuristics and may provide recommendations to improve the interface's usability. This evaluation can be conducted at various stages of the design process, ranging from early wireframes to the final product.

Heuristic evaluation offers a quick and cost-effective way to identify usability issues early on and provide feedback to improve the product's design [5]. This method can be useful in improving the usability of a product and ensuring a positive user experience.

The evaluation of usability often involves the use of heuristic principles or guidelines that are established based on research and experience. Ten commonly used heuristic principles for evaluating usability are [6]:

- 1) Visibility of system status: The system should provide appropriate feedback within a reasonable amount of time to keep the user informed about what is happening.
- 2) Match between system and the real world: The system should use words, phrases, and concepts that are familiar to the user, and follow real-world conventions.
- 3) User control and freedom: Users should be provided with options to easily reverse actions and resume tasks.
- 4) Consistency and standards: The system should follow platform conventions, and avoid confusing users with different words, situations, or actions.

- 5) Error prevention: The system should be designed to prevent serious user mistakes by putting up constraints or barriers.
- 6) Recognition rather than recall: The user should not have to remember information from one part of the interface to another, and objects, actions, and options should be visible.
- 7) Flexibility and efficiency of use: The system should cater to both experienced and novice users, and allow the user to tailor frequent actions and automate repetitive tasks.
- 8) Aesthetic and minimalist design: The interface should prioritize important information, and avoid containing irrelevant or rarely needed information.
- 9) Help users recognize, diagnose, and recover from errors: Error messages should be expressed in plain language and clearly indicate the problem and steps to resolve it.
- 10) Help and documentation: The system should include documentation to help users understand features and capabilities, but the system should be designed in such a way that documentation is not needed to complete tasks.

The use of these heuristic principles can help designers and evaluators identify usability issues and improve the overall user experience of a software application [4].

## B. Procedures

The expert evaluation involved the expert assessment of three software applications developed by Delta3 GmbH Software, namely the iOS delta3 app, the mac OS application Editor, and the web-based operations portal. The evaluation was conducted on devices running the latest operating systems, which included the iPhone 13 with iOS 16.1, the iPad Air 2022 with iPadOS 16.1, and the MacBook Air M1 with macOS 13.1.

The iOS app and macOS application were configured using the provided server settings, and a user account was created for each UI (App/Player, Editor, Operations) to facilitate the heuristic evaluation process. During the evaluation, the expert identified a number of strengths and problems associated with each of the software applications.

## C. Results

### 1) App/Player:

Description	Image	Location
The settings and advanced option	refer to Appendix B Fig.4	Settings

TABLE I: Usability strengths for App/Player

Description	Image	Location	Violation	Severity Rating
Unable to see complete name of the Project or solution	refer to Appendix B Fig.5	Your worksheet	Heuristic 1	Medium
Need to login for each new session	refer to Appendix B Fig.6	Login	Heuristic 7	Low
Unable to understand project progress and worksheet progress	refer to Appendix B Fig.7	Your worksheet	Heuristic 10	Low
Unable to see text in dark mode	refer to Appendix B Fig.8	worksheet	Heuristic 1	High

TABLE II: Usability problems for App/Player

During the evaluation of the App/Player, the expert identified a range of strengths and problems. One notable strength of the application was its user-friendly settings and advanced options, which were conveniently located in the settings option, refer to TABLE I. However, the expert also observed several issues that detracted from the overall user experience. Specifically, the first problem related to the absence of the project name in the user's worksheet, which was considered a violation of heuristic principle 1 and assigned a medium severity, refer to TABLE II. The second problem involved the lack of auto-login functionality on the login page, which violated heuristic principle 7 and was deemed to have a low severity, refer to TABLE II. The third problem related to the difficulty in understanding project progress, which was situated in the user's worksheet and violated heuristic principle 10, with low severity. Lastly, the fourth problem was the invisibility of the text in dark mode, which violated heuristic principle 1 and was assigned a high severity, refer to TABLE II.

## 2) Editor:

Description	Image	Location
Minimalistic design (Settings window)	refer to Appendix B Fig.9	Settings
The option to add custom grid	refer to Appendix B Fig.10	Add node
The paste and match option	refer to Appendix B Fig.11	Edit menu

TABLE III: Usability strengths for Editor

Description	Image	Location	Violation	Severity Rating
Missing UNDO feature (Ctrl + z)	refer to Appendix B Fig.12	Editor	Heuristic 3	Medium
Horizontal scroll always zoom's out	refer to Appendix B Fig.13	Editor	Heuristic 7	Low
Horizontal and Vertical Scroll Bar	refer to Appendix B Fig.14	Editor	Heuristic 7	Low
Icons without label	refer to Appendix B Fig.15	Editor	Heuristic 4	High
Different nomenclature in documents and current version	refer to Appendix B Fig.15	Editor	Heuristic 10	High

TABLE IV: Usability problems for Editor

The evaluation of the Editor software yielded a number of strengths and problems, refer to TABLE III. One notable strength of the software was its minimalistic design, particularly in the Settings window, which was conveniently located in the settings option. The software also had the ability to add a custom grid, which was available in the toolbox, and a paste and match option in the context menu, both of which were identified as strengths, refer to TABLE III.

However, several issues were identified during the evaluation of the Editor software. The first problem was the absence of the UNDO keyboard shortcut (Ctrl + z), which was deemed a violation of heuristic principle 3 and assigned a medium severity, refer to TABLE IV. The second problem was the horizontal scroll, which functioned as a zoom option and violated heuristic principle 7, with low severity. The third problem was the absence of both horizontal and vertical scroll bars, violating heuristic principle 7 and also assigned low severity, refer to TABLE IV. The fourth problem was the presence of icons without labels in the left toolbox, which violated heuristic principle 4 and had a high severity. Finally, the fifth problem was the observation of different nomenclature in the documentation and current software version, which violated heuristic principle 10 and had a high severity, refer to TABLE IV.

## 3) Operations:

Description	Image	Location
The simple user setting	refer to Appendix B Fig.17	Profile
Overall minimalist and easy to understand UI	refer to Appendix B Fig.18	Dashboard

TABLE V: Usability strengths for Operations

Description	Image	Location	Violation	Severity Rating
Difficult to understand the concept of project and worksheet	refer to Appendix B Fig.19	Projects	Heuristic 2	Low
HAdd Worksheets button is disabled but looks like enabled	refer to Appendix B Fig.20	Projects	Heuristic 2	Low
Allows to save worksheet without Assigned to, it results in no visible project on the player.	refer to Appendix B Fig.21	Add Worksheet	Heuristic 5	High
Dark mode does not work	refer to Appendix B Fig.22	Settings	Heuristic 7	Low
If there is any error, the back button does not work.	refer to Appendix B Fig.22	Operator	Heuristic 6	Low

TABLE VI: Usability problems for Operations

During the heuristic evaluation of the Operations web-page, multiple strengths and problems were identified, refer to TABLE V. The evaluation revealed that the web-page had a simple user settings interface located in the profile section, and an overall minimalistic and user-friendly dashboard interface, which were identified as strengths.

However, several problems were observed during the evaluation of the Operations web-page, refer to TABLE VI. The first problem was the difficulty in understanding the concept of projects and worksheets, which was located in the projects option and violated heuristic principle 2 with low severity. The second problem was that the "Add worksheet" button appeared enabled but was actually disabled, violating heuristic principle 2 with low severity, refer to TABLE VI. The third problem was that the software allowed users to save worksheets without assigning them to a project, resulting in no visible projects on the player app, violating heuristic principle 5 with high severity. The fourth problem was that the dark mode feature did not work, which was located in the settings and violated heuristic principle 7 with low severity. The fifth problem was that if an error occurred, users could not go back to the previous page, violating heuristic principle 6 with low severity, refer to TABLE VI.

The above observations demonstrate the significance of expert evaluation in identifying usability issues in software applications. By identifying these problems, software developers and designers can make improvements that enhance the overall user experience and usability of the software. While additional problems were identified during the evaluation, only the top five are listed in the above table.

### III. USABILITY TESTS

#### A. *Methods*

##### 1) *User study:*

The study comprised seven distinct in-person tasks that were performed on an iPad Air 11". These tasks included logging in, changing language and theme settings, retrieving specific information, and performing a crane origami.

The primary aim of the study was to evaluate the usability of the iOS app/player from the users' perspective. To achieve this goal, the study employed in-person tasks to observe user behavior and gather feedback.

The study sought to identify any usability issues or areas for improvement in the app/player. These findings could then be utilized to enhance the app/player in future updates.

##### 2) *User experience questionnaire:*

The User Experience Questionnaire is a standardized survey-based method used to evaluate the subjective experience of a user when interacting with a product, system, or service [7]. The UEQ questionnaire is available online at [www.ueq-online.org](http://www.ueq-online.org), and it is widely used in both the software industry and academia to assess the user experience of software applications.

The UEQ comprises 26 items that are designed to measure six key dimensions of user experience: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty [8]. Each of these dimensions is assessed by four items, resulting in a total of 24 items, with an additional two items used to evaluate overall user experience and the user's willingness to recommend the software to others [9].

The six dimensions of user experience that are evaluated by the UEQ are defined as follows [10]:

- 1) **Attractiveness:** Refers to the aesthetic appeal of the software and its visual design. This includes factors such as the use of color, typography, and graphic elements.
- 2) **Perspicuity:** Refers to the clarity and simplicity of the software's interface and its ease of use. This includes factors such as the organization of information and the consistency of design elements.
- 3) **Efficiency:** Refers to the ease and speed with which users can accomplish their goals using the software. This includes factors such as the speed of response and the availability of shortcuts.
- 4) **Dependability:** Refers to the reliability and consistency of the software's performance. This includes factors such as error handling and system stability.
- 5) **Stimulation:** Refers to the degree to which the software engages and motivates users. This includes factors such as the use of multimedia elements and interactive features.
- 6) **Novelty:** Refers to the degree to which the software offers new or innovative features or functions. This includes factors such as the use of cutting-edge technology and unique design elements.

The UEQ questionnaire utilizes a 7-point Likert scale to rate each of its 27 items, with 1 representing strong disagreement and 7 indicating strong agreement. The two additional items used to evaluate overall user experience and willingness to recommend the software to others are also rated on the same 7-point scale [9].

The UEQ can be administered either online or in person, and is commonly used in conjunction with other usability testing techniques, such as heuristic evaluation or usability testing. The results of the UEQ can be leveraged to pinpoint areas of strength and weakness in the software's user experience, which can inform future design decisions. The UEQ is demonstrated to possess high levels of reliability and validity in measuring user experience across a range of software contexts [8].

Through the use of the UEQ to evaluate software usability, designers and developers can gain valuable insights into the user experience, identify opportunities for improvement, and make informed design decisions based on data. This approach can result in software applications that are more user-friendly, efficient, and effective, ultimately leading to higher levels of user satisfaction and improved business outcomes [7].

## *B. Procedures*

The current study reports on a usability survey of the Delta3 app/player, conducted with a sample of five participants who were purposefully selected for their lack of prior experience with the software. The survey was conducted in two languages, English and German, indicating a global focus. Participants were drawn from diverse backgrounds, including students, working professionals, and a graphics designer, with ages ranging from 24 to 32 years old. By selecting participants with no prior experience, the study aimed to reduce the influence of prior knowledge or opinions on the results, ensuring the objectivity and reliability of the survey. Furthermore, a task-based study was conducted with the same group of participants, who were required to complete a set of tasks, including creating an origami model.

Following the completion of a user study, an online user experience questionnaire survey was administered to record users' opinions. In addition to collecting responses to the 26-item User Experience Questionnaire, information on device type, user age, identified problems and strengths, and recommendations for improvement were also documented.

## *C. Results*

### *1) User observations:*

User 3V5G0, a 25-year-old student of Applied Entrepreneurship, completed the first three tasks quickly, without any hesitation or queries, indicating familiarity with such applications. However, she faced some difficulty while performing task 4 and required intervention to complete it in approximately 20 minutes. She completed the remaining tasks in 3-4 minutes.

User L5YVB, also a 25-year-old student of Applied Entrepreneurship, completed the first three tasks without any queries in a reasonable time of 5 minutes. While performing task 4, she appeared stressed but completed it in 10 minutes with minimal intervention. The remaining tasks were completed in the next 5 minutes.

User S94H6, a 32-year-old vocationally trained working professional, faced some difficulty with tasks 2 and 3, taking 6-7 minutes to complete them. Task 4 required a significant amount of intervention, and after watching videos multiple times, he was able to complete it in 18 minutes. He completed the remaining tasks in 5 minutes but appeared stressed throughout the process.

User K8E5Y, a 24-year-old professional graphics designer, completed the first three tasks in a quick time of 2 minutes and completed task 4 in just 6 minutes due to prior knowledge of origami. He was able to perform tasks 5, 6, and 7 in the next 4 minutes, appeared confident during the process, and provided valuable feedback.

User MW1JY, a 27-year-old financial advisor, completed the first three tasks in 5 minutes and needed multiple interventions during task 4. He was able to complete the task in 14 minutes after watching videos multiple times and took 7 minutes to complete the remaining tasks. He appeared stressed while performing task 4.

The study results revealed that certain participants displayed familiarity with the application and performed tasks with ease, while others required assistance and faced difficulties during certain tasks. These findings can be utilized to enhance the design and functionality of similar applications in the future.

User Task Rating									
User	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Strethths	Problems
3V5G0	1	1	1	2	3	1	3	Clean, clear UI, plenty of info for completing a task	Loading, dark mode text, workflows page layout
L5YVB	2	2	1	2	1	2	1	Good design, settings, easy	Order, settings layout, text overlap, icons, pictures
S94H6	1	2	1	3	1	2	2	User friendly, Colors, organisation	Difficult to navigate, font was small
K8E5Y	1	2	3	1	1	2	2	Simplistic and clean, Logging loading screen, Task progress indication, Snappy and responsive, Consistant design	Nain Menu-scroll bar, list or icons, Language settings should be in advanced settings, Setting window should be in center and not in the bottom, Add icons to settings options, Text overlap in main page, Worksheet - looks confused, not proper layout, Worksheet metadata should be at one place like priority and due date, Add space between progress bars, Pick workflow button name can be changed to instruction, Video keeps playing in background if you press next arrow
MW1JY	1	2	1	2	1	2	1	Intuvative, inportant points visible directly, Sooth working of videos	Workflow button name, language selection, datk mode selection settings can be organised

TABLE VII: User task rating

Each participant rated the difficulty level of each task on a scale of 1 to 5 (1 being easy and 5 being difficult) as presented in the table. Moreover, the participants provided feedback regarding the strengths and problems of the application, which are listed in the TABLE VII. The mean and standard deviation were calculated for each task to determine the level of difficulty. Task 4 was found to be the most difficult, with a mean of 2.0 and a standard deviation of 1.0, refer to TABLE VII. Positive feedback included the software's clean and clear UI, good design, user-friendly colors and organization, as well as intuitive and smooth working of videos. Negative feedback included issues with loading times, dark mode text, problems with workflows page layout, settings layout, text overlap, icons, pictures, difficult navigation, small font, list or icons in the main menu scroll bar, language settings not in advanced settings, setting window in the bottom rather than the center, text overlap in the main page, confused worksheet layout, improper layout of worksheet metadata, lack of space between progress bars, and issues with the video continuing to play in the background if the next arrow is pressed. In conclusion, while the software has some positive features, there are also several areas that need improvement, particularly regarding loading times, layout, and settings, refer to TABLE VII.

## 2) Mean value per item:

Based on the UEQ mean value per item, refer to Appendix A TABLE IX, the following observations can be made:

**Attractiveness:** The mean score for the attractiveness scale is 2.0, indicating moderate attractiveness, refer to Appendix A TABLE IX. However, item 1 "annoying" received a relatively low score of 1.8, suggesting that certain aspects of the product may be annoying to users, refer to Fig.1.

**Perspiciuity:** The mean score for the perspicuity scale is 1.8, indicating somewhat understandable. However, item 2 "not understandable" received a relatively low score of 2.2, indicating that some aspects of the product may be difficult to understand, refer to Fig.1. Additionally, item 13 "complicated" received a low score of 1.2, further indicating issues with ease of use, refer to Appendix A TABLE IX.

**Novelty:** The mean score for the novelty scale is 0.8, indicating low perceived novelty or creativity. Item 3 "creative" received a low score of 0.6, suggesting a lack of innovation, refer to Appendix A TABLE IX. Items 10 "inventive" and 15 "leading edge" received negative scores, indicating that users did not find these aspects of the product to be innovative or unique, refer to Appendix A TABLE IX.



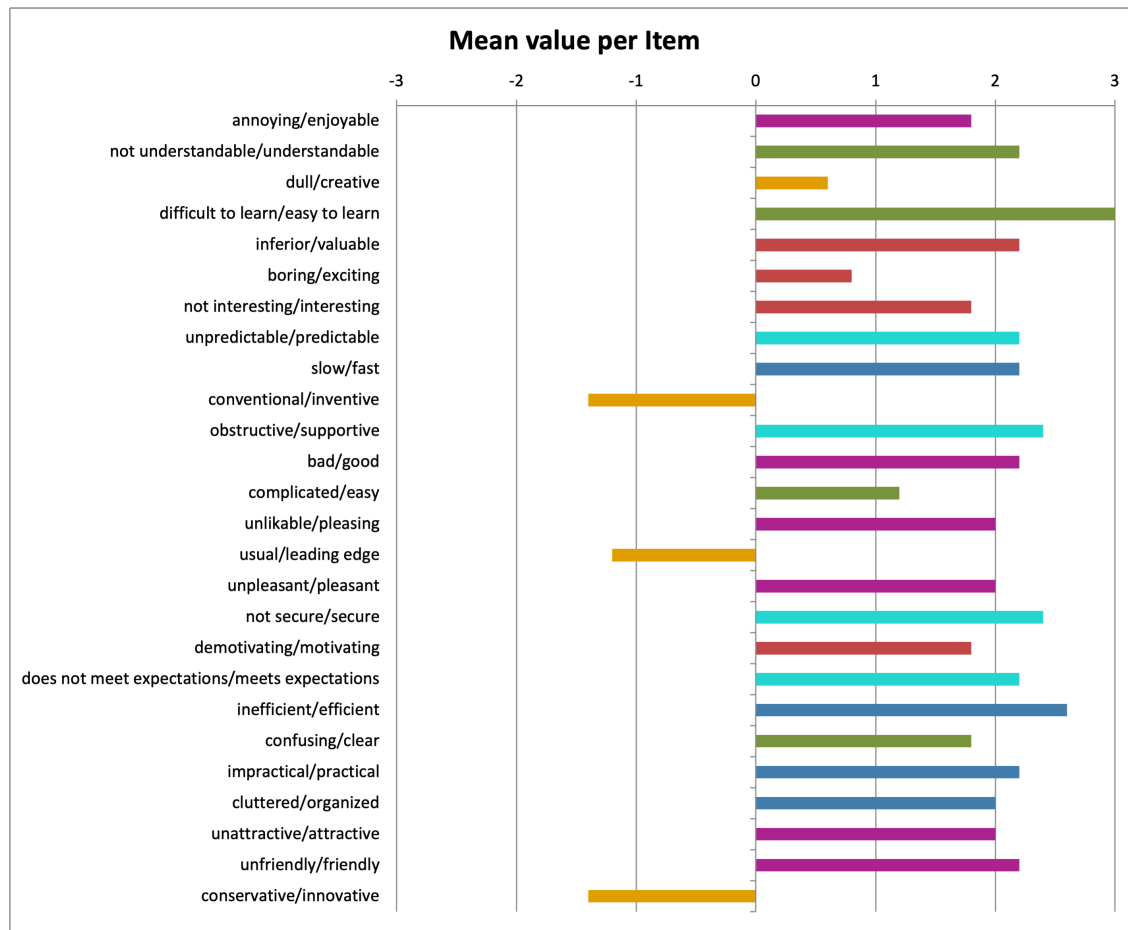


Fig. 1: Mean value per item

**Stimulation:** The mean score for the stimulation scale is 1.7, indicating moderate stimulation, refer to Appendix A TABLE IX. Item 5 "valuable" received a relatively high score of 2.2, indicating perceived value. However, items 6 "boring" and 18 "demotivating" received low scores, suggesting unengaging aspects, refer to Fig.1.

**Dependability:** The mean score for the dependability scale is 2.2, indicating relatively dependable. Items 8 "unpredictable", 11 "obstructive", and 19 "does not meet expectations" received scores below the mean, indicating issues with reliability or consistency, refer to Appendix A TABLE IX.

**Efficiency:** The mean score for the efficiency scale is 2.2, indicating somewhat efficient, refer to Appendix A TABLE IX. Item 9 "fast" received a relatively high score of 2.2, indicating perceived speed. However, items 20 "inefficient", 22 "impractical", and 23 "cluttered" received scores below the mean, indicating usability and efficiency issues, refer to Fig.1.

These UEQ results suggest that while the product was generally moderately attractive and dependable, there are issues with ease of use, innovation, and efficiency. Certain aspects of the product were also perceived as unengaging or demotivating. These insights can guide product designers and developers in identifying areas for improvement and prioritizing changes to improve user experience.

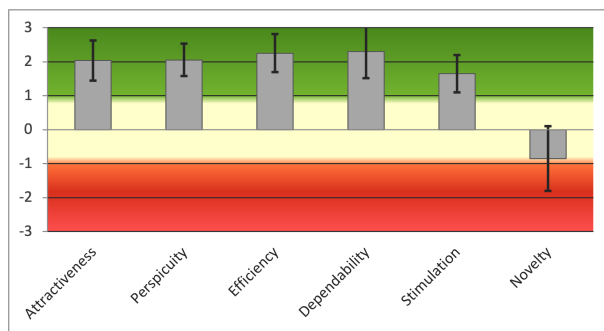
### 3) UEQ scales:

The study investigated the means, variances, and standard deviations of the ratings for each item, as well as the means and variances for each of the six UEQ scales: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. The findings showed that the highest mean score was for Efficiency (2.250), followed by Dependability (2.300), and then Attractiveness and Perspicuity (both at 2.033 and 2.050, respectively). The lowest mean score was for Stimulation (1.650), and Novelty had a negative mean score (-0.850), indicating a lack of novelty in the product or service, refer to Fig.2a and Appendix A TABLE X.

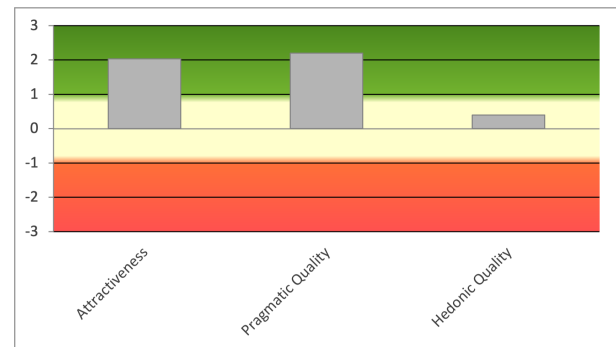
The variances for each scale indicated the amount of variation in the responses. Dependability had the highest variance (0.79), suggesting mixed experiences with the product or service's dependability, refer to Appendix A TABLE X. Novelty had the highest variance (1.18), indicating highly varied experiences with the product or service's novelty, refer to Fig.2a.

Analyzing the individual items, Item 13 had the highest variance (2.4), implying highly varied experiences with the product or service's complexity. Item 10 had a negative mean score (-1.4), indicating that users did not find the product or service inventive, refer to Appendix A TABLE X. Item 26 also had a negative mean score (-1.4), implying that users did not find the product or service innovative, refer to Appendix A TABLE X.

The results suggested that users found the product or service to be efficient and dependable but not stimulating or novel. There were mixed experiences with attractiveness and perspicuity.



(a) UEQ Scales



(b) Pragmatic and Hedonic Quality

Fig. 2: Results UEQ Scales and Pragmatic & Hedonic Quality

#### 4) Pragmatic & Hedonic Quality:

Based on the UEQ pragmatic and hedonic quality dimensions, the overall attractiveness score of the user experience was 2.03, refer to Appendix A TABLE XI. This score suggests that the participants rated the user experience slightly above average in terms of its overall attractiveness, refer to Fig.2b.

The pragmatic quality score, which is a composite of Perspicuity (clarity), Efficiency, and Dependability, had a mean score of 2.20, refer to Appendix A TABLE XI. This score indicates that the participants found the user experience to be slightly above average in terms of task-related quality aspects, refer to Fig.2b.

The hedonic quality score, which is a composite of Stimulation and Novelty, had a mean score of 0.40. This score suggests that the participants found the user experience slightly below average in terms of non-task-related quality aspects, refer to Appendix A TABLE XI.

Of particular note is the negative mean score for Novelty (-0.85), which indicates that participants found the user experience less novel than expected, refer to Fig.2b.

The results suggest that the user experience is generally perceived as slightly above average in terms of pragmatic quality, but slightly below average in terms of hedonic quality, with low scores specifically in terms of Novelty.

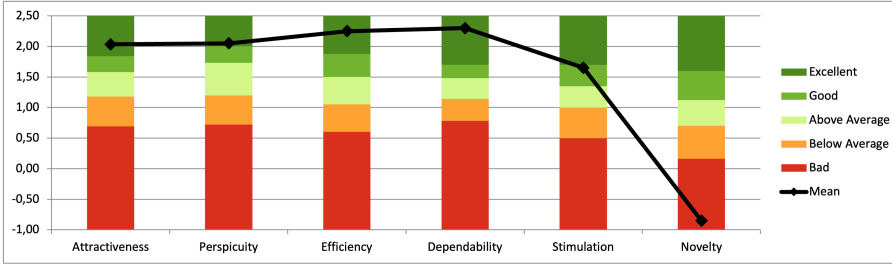
#### 5) Benchmark:

Based on the benchmark results, the evaluated product has demonstrated impressive performance in terms of attractiveness, perspicuity, efficiency, and dependability, scoring among the top 10% compared to other products. This indicates that users perceive the product as visually appealing, easy to understand, efficient to use, and reliable.

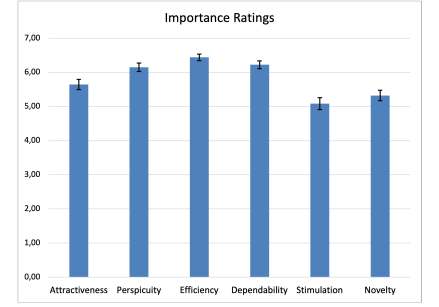
However, the product's performance in terms of stimulation falls among the bottom 75% of results, suggesting that users may not find the product very engaging or exciting, which could affect their overall satisfaction, refer to Fig.3a.

Additionally, the product's novelty score is among the bottom 25% of results, indicating that it may not be particularly innovative or unique compared to other products, refer to Fig.3a. While novelty may not be essential for all products, it can be crucial in highly competitive markets or products that require frequent updates to maintain user engagement.

The benchmark shows that the evaluated product excels in pragmatic quality, but there is room for improvement in hedonic quality and novelty.



(a) UEQ Benchmark



(b) KPI Calculation

Fig. 3: Benchmark and KPI Calculation

#### 6) KPI calculation:

The descriptive statistics of the UEQ scales, including mean, standard deviation, sample size, and confidence interval.

To derive a Key Performance Indicator (KPI), a formula is necessary to combine the individual scale scores. A possible approach is to compute the arithmetic mean of all UEQ scale scores [9].

With this formula, the resulting KPI is [9]:

$$\text{KPI} = (5.64 + 6.15 + 6.44 + 6.22 + 5.09 + 5.32) / 6 = 5.67, \text{ refer to Fig.3b}$$

This value represents the overall user experience score for the evaluated product, ranging from 0 to 8.

The confidence intervals provide information about the precision of the estimated mean values. For instance, the attractiveness scale's estimated mean score is 5.64, with a 95% confidence interval of [5.49, 5.79]. This indicates that there is a 95% probability that the actual mean score of the population falls within this interval. Similarly, we can interpret the confidence intervals of the other UEQ scales.

#### 7) Sample size:

Condition	Attractiveness	Perspicuity	Efficiency	Dependability	Stimulation	Novelty
Precision=0.5,Err.Prob.=0.1	5	3	4	9	4	13
Precision=0.5,Err.Prob.=0.05	7	5	6	12	6	18
Precision=0.5,Err.Prob.=0.01	12	8	11	21	10	31
Precision=0.25,Err.Prob.=0.1	20	13	18	35	17	51
Precision=0.25,Err.Prob.=0.05	28	18	25	49	24	72
Precision=0.25,Err.Prob.=0.01	48	31	43	85	42	125
Precision=0.1,Err.Prob.=0.1	123	80	111	216	107	320
Precision=0.1,Err.Prob.=0.05	173	113	156	305	151	451
Precision=0.1,Err.Prob.=0.01	300	196	270	528	262	782

TABLE VIII: Sample Size

The present study has calculated various sample sizes for different levels of precision and error probability for each condition of the User Experience Questionnaire. The sample size estimation has been based on the assumption of a normal distribution with a standard deviation of 1, refer to TABLE VIII.

For instance, taking the first row as an example, with a precision of 0.5 and an error probability of 0.1, the sample size for the Attractiveness condition is 5. This implies that to estimate the mean score for Attractiveness in this condition with a precision of 0.5 and an error probability of 0.1, researchers must have a sample size of at least 5, refer to TABLE VIII.

Upon examining the table, we observe that the sample size increases as the precision and error probability become smaller. This trend occurs because higher precision and lower error probability require a larger amount of data to achieve statistical significance, refer to TABLE VIII.

The TABLE VIII provides valuable insights for researchers interested in conducting studies utilizing the UEQ. By following these sample sizes as a guide, researchers can ensure that their studies have sufficient statistical power to detect meaningful differences between conditions.

#### IV. DISCUSSION

The expert evaluation was conducted by an expert on three software applications developed by Delta3 Gmbh Software, namely the iOS delta3 app, the macOS application Editor, and the web-based operations portal. The evaluation identified a range of strengths and problems associated with each of the software applications. The observations demonstrate the importance of heuristic evaluation in identifying usability issues in software applications and guiding software developers and designers in making improvements that enhance the overall user experience and usability of the software. Based on the UEQ benchmark results, the analyzed product has performed exceptionally well in all categories except for stimulation and novelty. Specifically, the attractiveness, perspicuity, efficiency, and dependability scales scored within the 10% best results, indicating a high level of user satisfaction with the product in terms of these attributes. However, the stimulation scale had a good rating, indicating that 10% of results were better, but 75% of results were worse, suggesting that there is some room for improvement in terms of providing a more engaging and stimulating user experience. Additionally, the novelty scale had a bad rating, suggesting that the product's novelty score is in the range of the 25% worst results. This indicates that the product may not be perceived as innovative or fresh by users, and there may be a need to introduce more novel and creative features to improve user experience.

The study tested the usability of the Delta3 app/player with five participants of diverse backgrounds. Some completed tasks with ease, while others faced difficulties. Task 4 was the most difficult. Participants provided feedback on the app's strengths and problems, including issues with loading times, layout, and settings. The study's findings can be used to improve similar applications in the future.

The UEQ KPI calculation provides a more detailed analysis of user experience by measuring mean scores, standard deviation, confidence intervals, and benchmark comparisons across all six UEQ scales. The results indicate that the product scored highest in the efficiency scale (mean=6.44), followed by the dependability scale (mean=6.22), and perspicuity scale (mean=6.15). These scores are all above the benchmark comparison, indicating that users were highly satisfied with these aspects of the product. The attractiveness scale had a mean score of 5.64, which is also above the benchmark comparison. However, this score had a relatively higher standard deviation (1.21), suggesting that user opinions were more diverse on this scale. The stimulation and novelty scales had lower mean scores of 5.09 and 5.32, respectively, which were also below the benchmark comparison. These results indicate that the product may need to improve in terms of providing a more stimulating and novel user experience.

Lastly, the Cronbach's alpha coefficient analysis provides insights into the reliability of the UEQ measurement instrument used in the study. The alpha coefficient score of 0.86 indicates that the UEQ is highly reliable in measuring user experience across all six scales. The inter-item correlations suggest that the items in each scale are positively correlated, indicating that they measure the same underlying construct. However, some items in the perspicuity and efficiency scales have a "DIV/0!" value, which could affect the reliability of these scales. The analysis also provides a confidence interval for the alpha coefficient, indicating that the true reliability of the instrument falls between 0.21 and 0.98 at a 5% level of significance.

Based on the analysis of the software usability results, it can be concluded that overall, the software has moderate usability, with an average UEQ score of 0.51. The highest rated aspect of usability was "Dependability" with a score of 0.76, while the lowest rated aspect was "Novelty" with a score of 0.31. The Cronbach's alpha coefficient for the UEQ was 0.86, indicating a high level of internal consistency.

However, it is important to note that the sample size for the UEQ was relatively small, with only 5 participants. This may limit the generalizability of the results and larger sample sizes should be considered in future usability testing.

Overall, the results suggest that while the software has some usability issues, it is generally acceptable for use. Further improvements could be made in the areas of stimulation and novelty to enhance user engagement and satisfaction.

## V. RECOMMENDATIONS

Based on the results of the expert evaluation, it is recommended that Delta3 Gmbh Software should address the identified problems in the iOS delta3 app, the macOS application Editor, and the web-based operations portal, in order to improve their usability and enhance the overall user experience. These problems should be addressed by following established usability principles and guidelines, and making improvements to the software's design and functionality.

In terms of the UEQ benchmark results, it is recommended that the company should focus on improving the stimulation and novelty scales of the analyzed product, in order to provide a more engaging and innovative user experience. This could be achieved by introducing more novel and creative features and enhancing the software's design to make it more visually appealing and stimulating. Additionally, it is recommended that Delta3 Gmbh Software address the issues identified in the study's feedback, such as loading times, layout, and settings, in order to improve the usability of similar applications in the future.

Furthermore, it is recommended that Delta3 Gmbh Software conduct further usability testing with a larger sample size to obtain more reliable and generalizable results. This will enable the company to obtain a more comprehensive understanding of users' experiences and improve the software's usability accordingly.

## ACKNOWLEDGMENT

I sincerely appreciate the help and guidance received from Prof. Dr. Dr. Dr. Carsten Röcker. I would especially want to thank Mario Heinz-Jakobs, M.Sc., for his continued assistance whenever it was required, and I would also like to express my sincere gratitude to all survey respondents who gave me insightful feedback.

APPENDIX A  
TABLES

Item	Mean	Variance	Std. Dev.	No.	Left	Right	Scale
1	1,8	0,2	0,4	5	annoying	enjoyable	Attractiveness
2	2,2	0,2	0,4	5	not	understandable	understandable Perspicuity
3	0,6	2,3	1,5	5	creative	dull	Novelty
4	3,0	0,0	0,0	5	easy to learn	difficult to learn	Perspicuity
5	2,2	0,2	0,4	5	valuable	inferior	Stimulation
6	0,8	1,7	1,3	5	boring	exciting	Stimulation
7	1,8	0,7	0,8	5	not interesting	interesting	Stimulation
8	2,2	1,2	1,1	5	unpredictable	predictable	Dependability
9	2,2	1,7	1,3	5	fast	slow	Efficiency
10	-1,4	2,3	1,5	5	inventive	conventional	Novelty
11	2,4	0,3	0,5	5	obstructive	supportive	Dependability
12	2,2	0,7	0,8	5	good	bad	Attractiveness
13	1,2	5,7	2,4	5	complicated	easy	Perspicuity
14	2,0	1,0	1,0	5	unlikable	pleasing	Attractiveness
15	-1,2	2,7	1,6	5	usual	leading edge	Novelty
16	2,0	0,5	0,7	5	unpleasant	pleasant	Attractiveness
17	2,4	0,8	0,9	5	secure	not secure	Dependability
18	1,8	0,2	0,4	5	motivating	demotivating	Stimulation
19	2,2	1,7	1,3	5	meets expectations	does not meet expectations	Dependability
20	2,6	0,3	0,5	5	inefficient	efficient	Efficiency
21	1,8	0,7	0,8	5	clear	confusing	Perspicuity
22	2,2	0,7	0,8	5	impractical	practical	Efficiency
23	2,0	1,0	1,0	5	organized	cluttered	Efficiency
24	2,0	1,5	1,2	5	attractive	unattractive	Attractiveness
25	2,2	0,7	0,8	5	friendly	unfriendly	Attractiveness
26	-1,4	2,3	1,5	5	conservative	innovative	Novelty

TABLE IX: Mean value per item

UEQ Scales (Mean and Variance)		
Attractiveness	2,033	0,45
Perspicuity	2,050	0,29
Efficiency	2,250	0,41
Dependability	2,300	0,79
Stimulation	1,650	0,39
Novelty	-0,850	1,18

TABLE X: UEQ Scales

Pragmatic and Hedonic Quality)	
Attractiveness	2,03
Pragmatic Quality	2,20
Hedonic Quality	0,40

TABLE XI: Pragmatic and Hedonic Quality

## APPENDIX B FIGURES

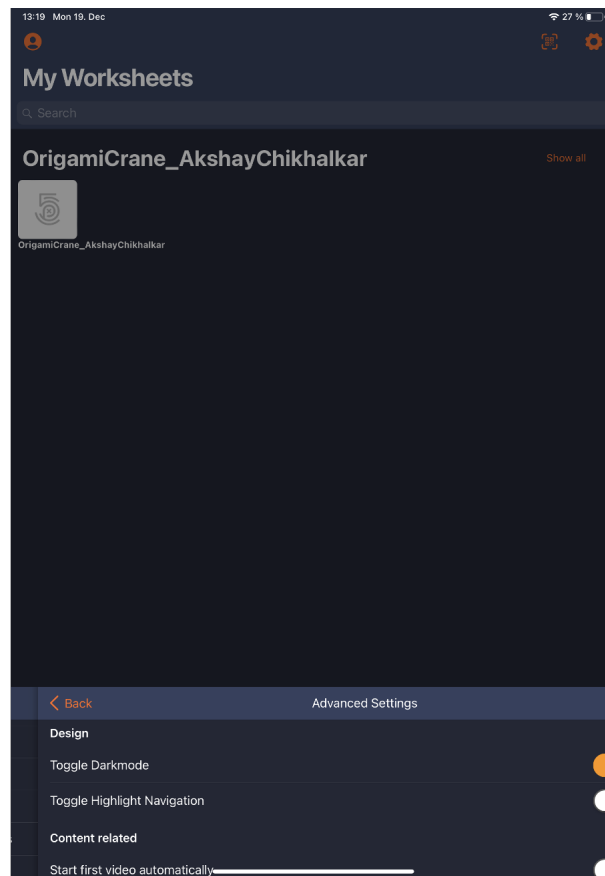


Fig. 4: App Strength 1

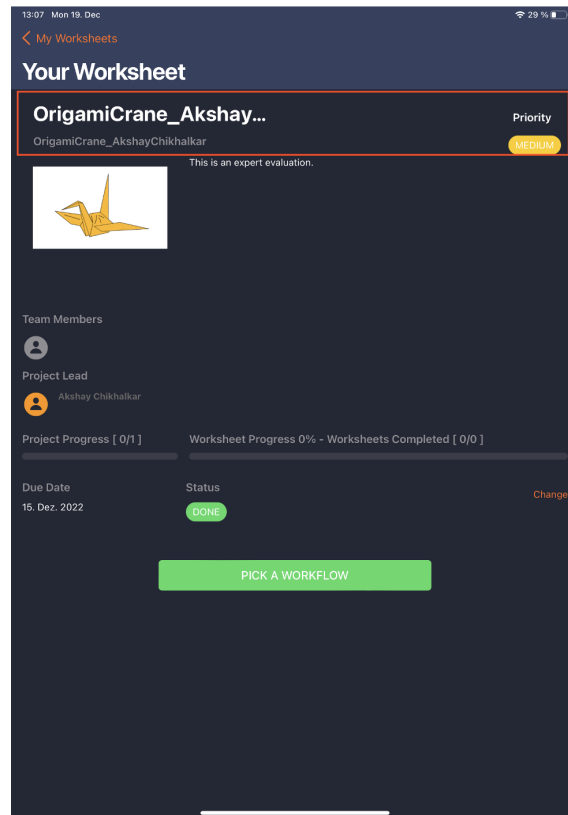


Fig. 5: App Problem 1



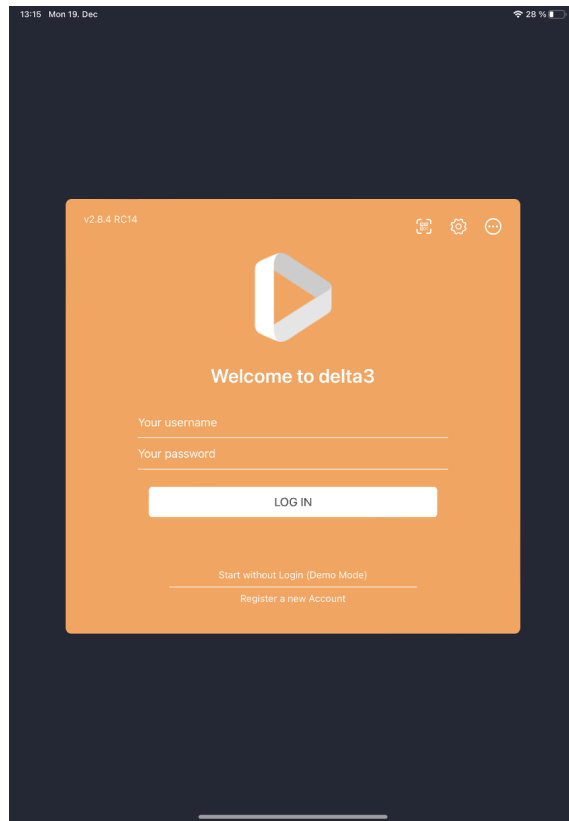


Fig. 6: App Problem 2

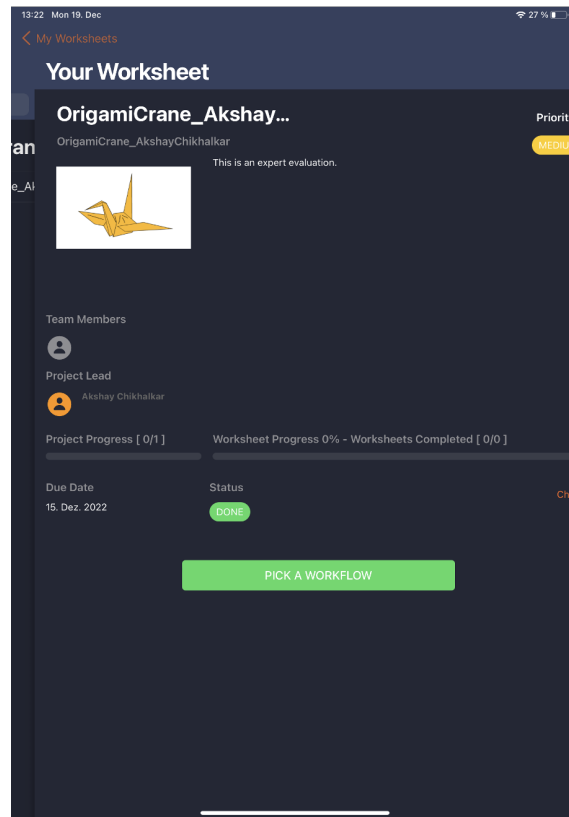


Fig. 7: App Problem 3

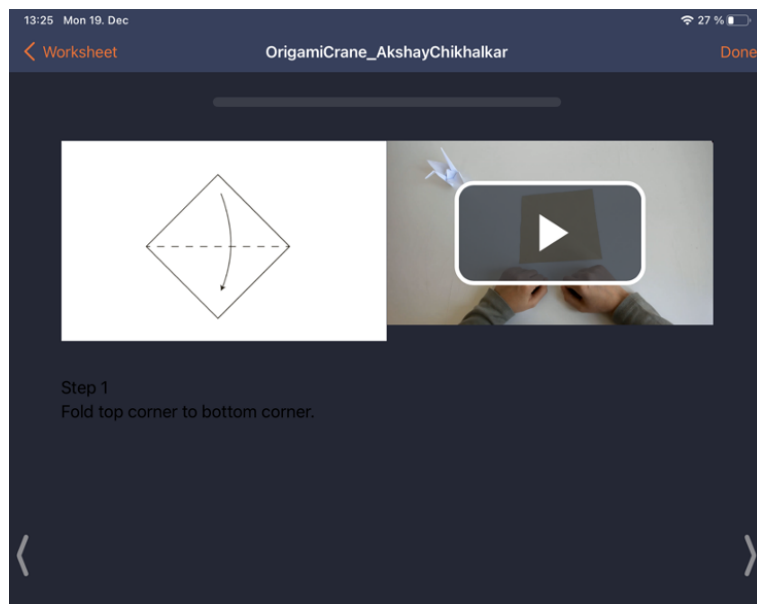


Fig. 8: App Problem 4

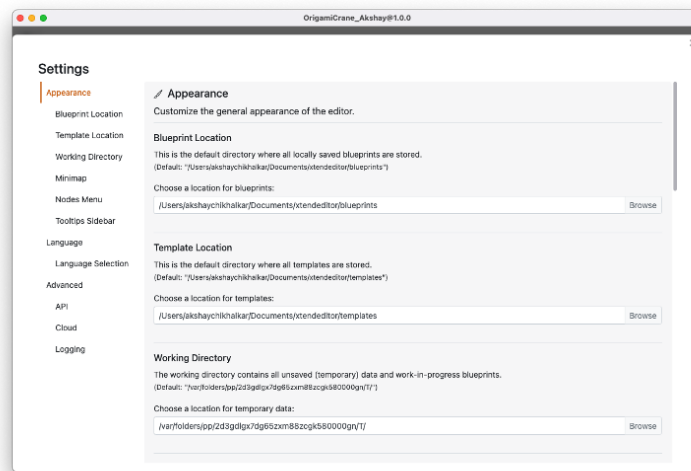


Fig. 9: Editor Strength 1

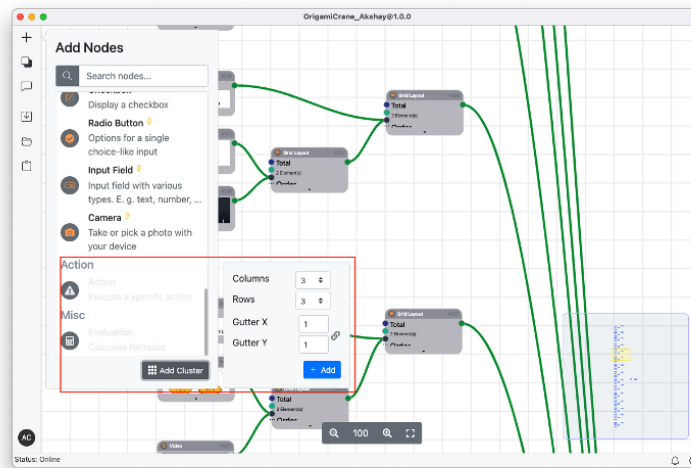


Fig. 10: Editor Strength 2

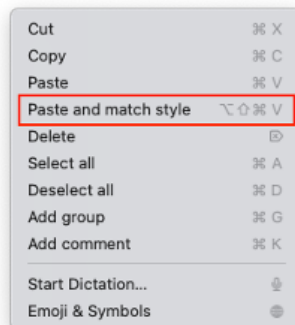


Fig. 11: Editor Strength 3

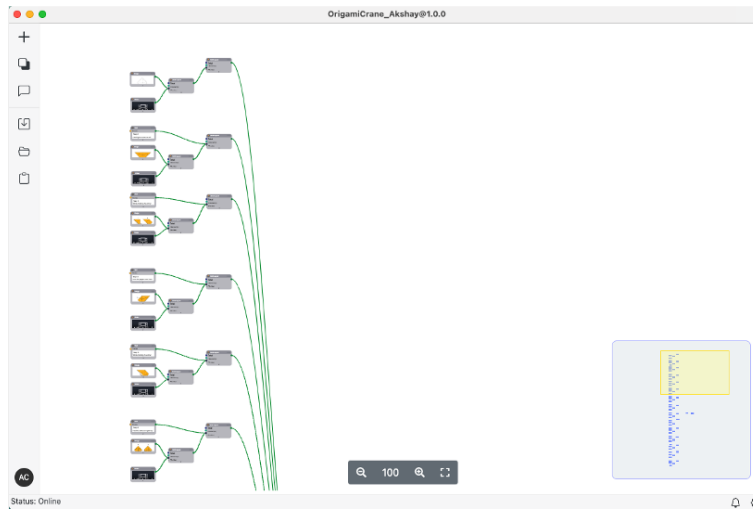


Fig. 12: Editor Problem 1

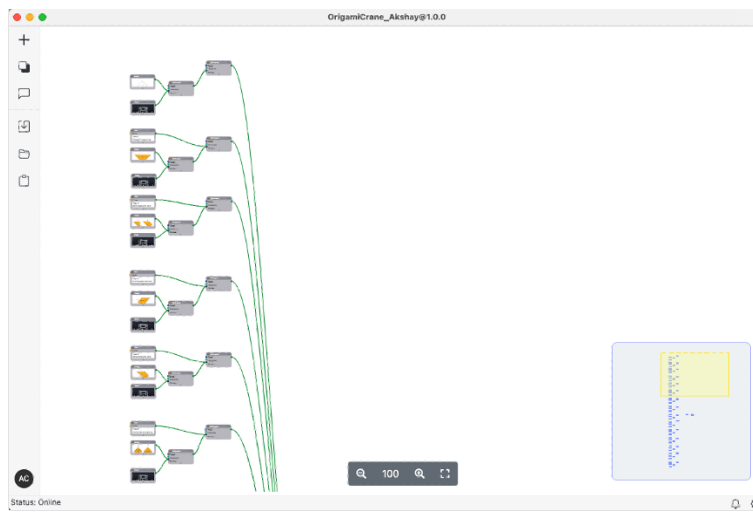


Fig. 13: Editor Problem 2

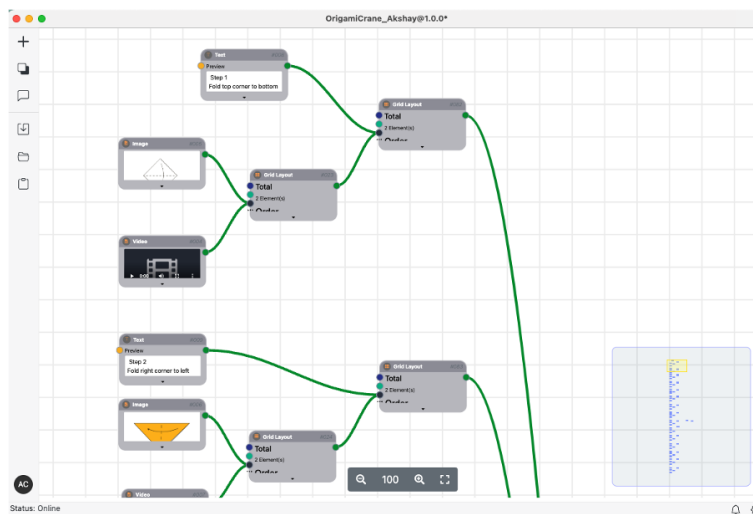


Fig. 14: Editor Problem 3



Fig. 15: Editor Problem 4

## Grid layout

### View:

#### Overview:

- Inputs
- Layout Output
- Edit grid view
- Number of columns
- Number of lines
- OK button



Fig. 16: Editor Problem 5

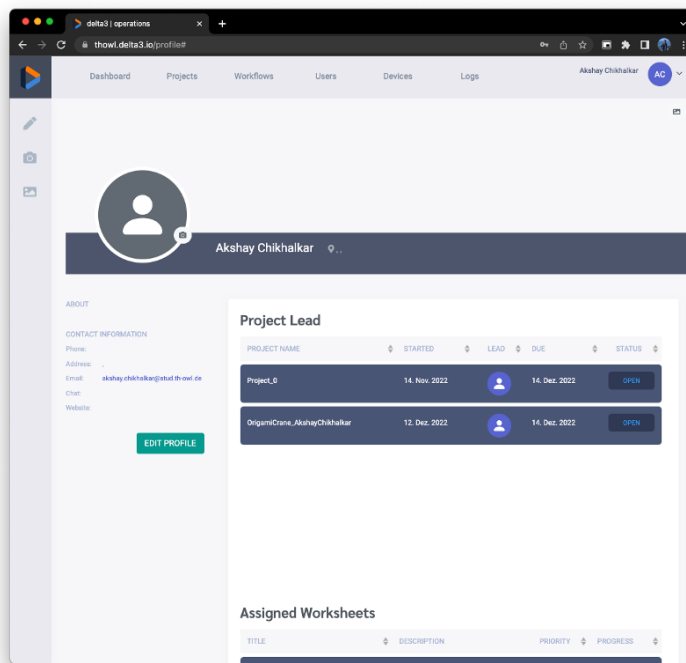


Fig. 17: Operation Strength 1

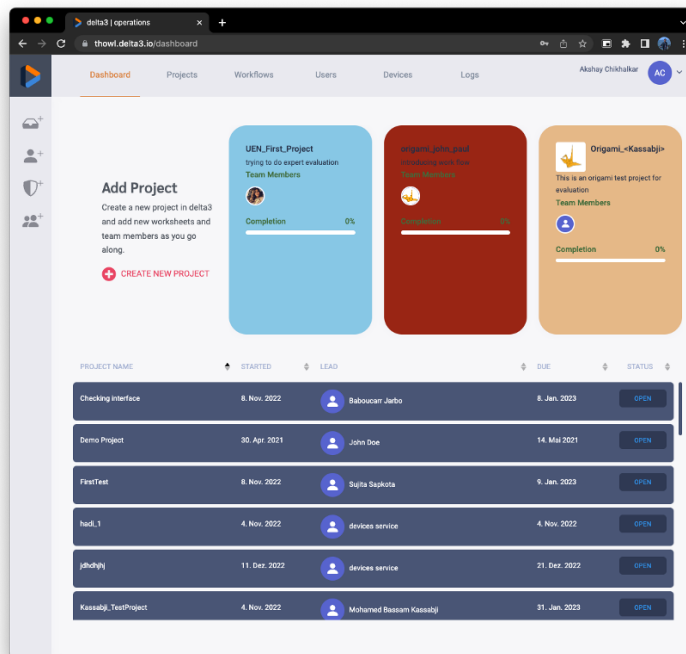


Fig. 18: Operation Strength 2

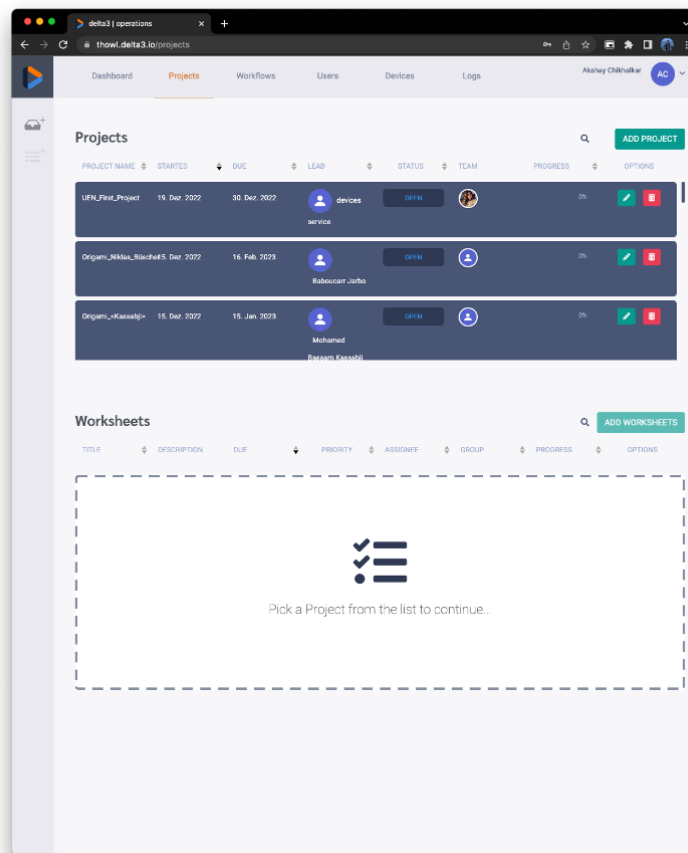


Fig. 19: Operation Problem 1

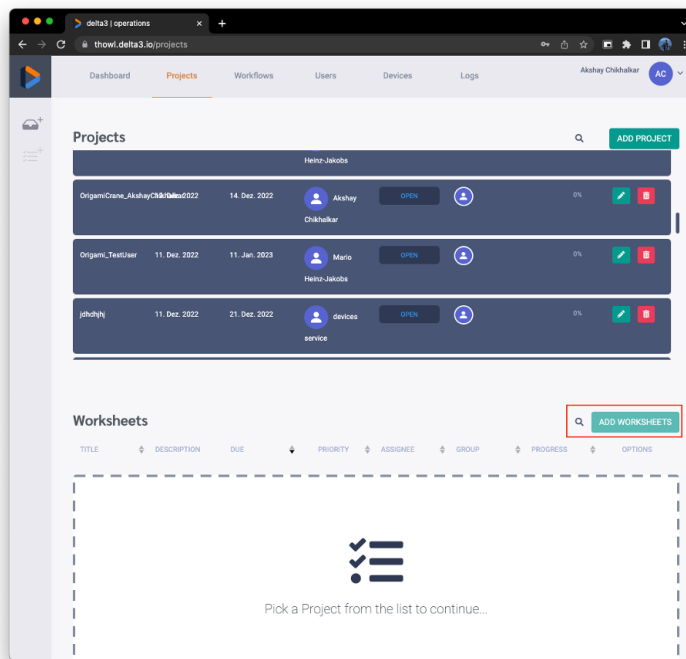


Fig. 20: Operation Problem 2

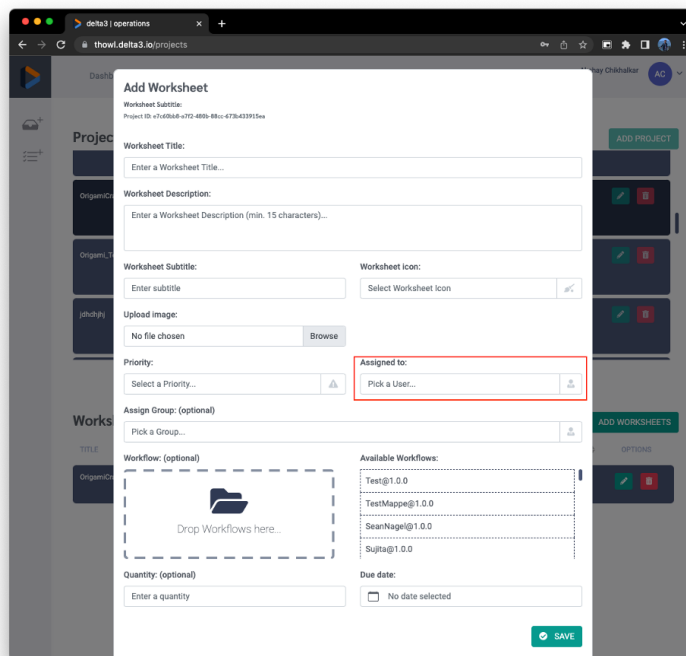


Fig. 21: Operation Problem 3



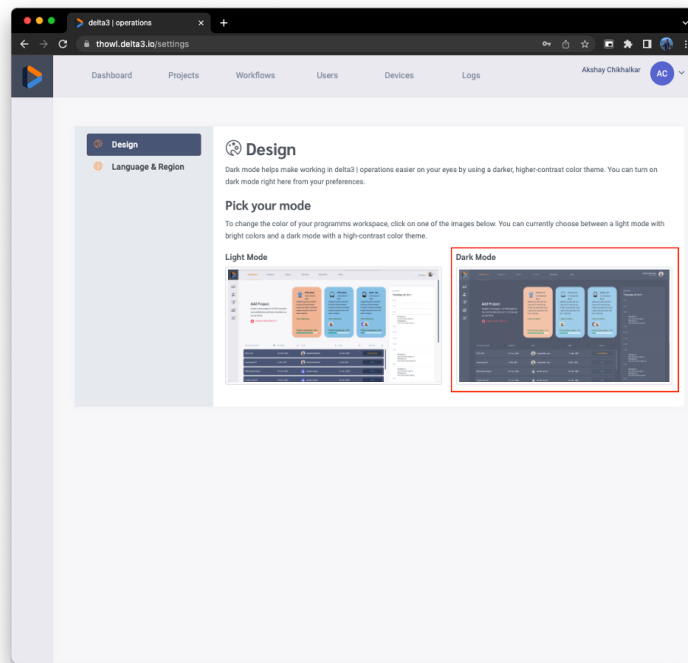


Fig. 22: Operation Problem 4

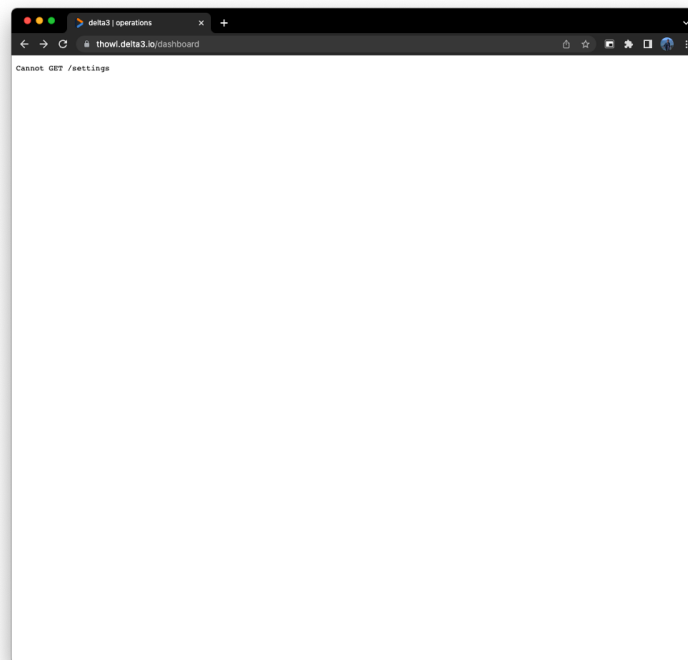


Fig. 23: Operation Problem 5

## REFERENCES

- [1] A. Dillon, "The evaluation of software usability," *Encycl Hum Factors Ergon*, 2001.
- [2] M. I. Berkman and D. Karahoca, "Re-assessing the usability metric for user experience (umux) scale," *Journal of Usability Studies*, vol. 11, no. 3, 2016.
- [3] K. Sagar and A. Saha, "A systematic review of software usability studies," *International Journal of Information Technology*, pp. 1–24, 2017.
- [4] J. Nielsen, "Heuristic evaluation," *Usability Inspection Mehods*, 1994.
- [5] R. Molich and J. Nielsen, "Improving a human-computer dialogue," *Communications of the ACM*, vol. 33, no. 3, pp. 338–348, 1990.
- [6] J. Nielsen, "Ten usability heuristics," 2005.
- [7] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.
- [8] M. Hassenzahl and N. Tractinsky, "User experience-a research agenda," *Behaviour & information technology*, vol. 25, no. 2, pp. 91–97, 2006.
- [9] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *HCI and Usability for Education and Work: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008, Graz, Austria, November 20-21, 2008. Proceedings 4*. Springer, 2008, pp. 63–76.
- [10] M. Hassenzahl, M. Burmester, and F. Koller, "Attrakdiff: A questionnaire to measure perceived hedonic and pragmatic qualityät," *Human & Computer 2003: Interaction in Motion*, pp. 187–196, 2003.