International Journal of Research in Interdisciplinary Studies
Volume 3, Issue 11, November 2025
https://www.ijris.com | ISSN (Online): 2584-1017

Evaluating Prototype Fidelity: Impacts on Cognitive Workload and Mental-Model Alignment in Flight-Booking Interfaces

Tanisi Das*
Undergraduate, School of Computer Science, Carleton University, Ottawa, Canada

Abstract—**Early-stage prototyping is essential for translating user requirements to functional design concepts. However, empirical evidence that investigates how prototype fidelity impacts cognitive workload and mental-model alignment within multi-step tasks is limited. This study examines three levels of fidelity (low-fidelity paper sketches, medium-fidelity clickable wireframes and high-fidelity static HTML pages) in a multi-step flight-booking scenario. Participants in this study included 25 undergraduate students who completed the moderate-complexity task workflow using a counterbalanced within-subjects design. The NASA-TLX was used to measure perceived cognitive workload and mental-model alignment was evaluated using a self-report Likert-scale questionnaire. Thus, the following results demonstrated a significant reduction in cognitive workload and increase in mental-model alignment with increase in fidelity. Further task performance analysis indicated that completion times were faster for medium-fidelity wireframes than for other conditions. Therefore, these findings present empirical guidance for selection of prototype fidelity which further demonstrates that medium-fidelity wireframes provide considerable cognitive and usability benefits along with reduced resources.**

Index Terms— **Cognitive Workload, Flight-Booking Interfaces, Human–Computer Interaction (HCI), Interface Complexity, Mental Models, NASA-TLX, Prototype Fidelity, Usability Testing, Wireframes.**

1. Introduction

Early-stage prototyping is crucial for transforming user requirements into a cohesive conceptual model during the development process of a proposed system. Prototypes allow designers to explore the behavior of the system - the core functions and services provided by the system, how they are related to one another and what information must be available to the users. These include low-fidelity prototypes such as paper sketches, medium-fidelity wireframes, and high-fidelity prototypes mock-ups. These prototyping forms establish a proper understanding of the client's requirements, user task and navigational workflows, the users' perception of the elements that make up the system - essentially, the final user goals and behaviors [1]. It is a crucial part of the development process and sets the foundation for not only the implementation phase but for how users expect to navigate the system and carry out tasks.

It is essential that prototypes match users' internal expectations - otherwise known as 'mental models' - the lack of which may result in excessive cognitive load and consequently issues in usability [2]. By evaluating the mental-model alignment and perceived workload during prototyping with different levels of fidelity, errors and oversights regarding the effort required for certain task workflows can be amended and refined - ensuring that the revised design aligns with user expectations and minimizes cognitive effort. These issues are illustrated within a moderate complexity workflow scenario such as booking flight tickets where related prototypes do not necessarily align with similar booking system scenarios. This would result in greater cognitive effort by the user as they navigate the system in order to interpret the controls, which diverts attention from the identification of usability issues. Assessing the mental-model alignment along with the perceived workload during the prototyping of the flight-booking task within low, medium and high-fidelity levels will allow the identification of steps within the workflow that may be the cause for misalignment or increased cognitive effort.

Failures in user interaction that highlight usability breakdowns highlight the discrepancy between the intended interface design and users' actual experiences. Design workflows in real life involve teams relying on either low-fidelity prototypes and creating quick paper sketches in order to conserve resources or skip to high-fidelity prototypes in order to present quick and polished product mock-ups and expedite the process to gain client approval. This often results in design teams neglecting to place emphasis on whether prototypes match user expectations and whether the prototypes are characterized by cognitive load heavy workflows [3]. In this case, usability issues may remain undetected until much later in the process. In spite of the usage of prototyping with different fidelities in the field, the body of quantitative literature that evaluates how prototype fidelity levels may affect the mental-model alignment and cognitive workload remains limited - as a result of which, designers do not possess evidence-based criteria that may assist in selecting appropriate fidelity prototypes. For instance, prior studies that investigate usability issues for paper and computer prototypes [4] do not necessarily observe how prototype types may affect users' mental models

*Corresponding author: tanisidas.21@gmail.com

or cognitive workload. Another study evaluating paper sketches and HTML front-end pages [5] failed to observe moderate-complexity workflows and perceived cognitive workload.

With the objective of addressing the gap within quantitative literature, the framework that supports the research questions is laid down. RQ1 examines how prototype fidelity level may affect participants' perceived cognitive workload (measured using the NASA-TLX questionnaire) during a moderate-complexity task - flight-booking while R2 investigates how prototype fidelity may affect mental-model alignment (measured using the self-reported questionnaire) during the same moderate-complexity task. These research questions address the impact on cognitive effort and the perceived workload that are observed to be largely unprobed areas of research. Accordingly, H1 predicts that higher levels of fidelity will result in lower perceived workload and hence lower NASA-TLX scores by the user, while H2 predicts that higher levels of fidelity will result in higher metal-model alignment scores.

This study conducts a quantitative comparison of the prototype fidelities - low, medium and high - within a task of moderate complexity, i.e., involving multiple steps of booking flight tickets. This involves evaluating user behavior as participants navigate the same task workflow within prototypes of different fidelity levels. In addition, the study will present qualitative findings related to balancing prototyping budgets against user experience using measures such as the NASA-TLX questionnaire and a mental-model alignment questionnaire. This assists in quantitative assessment of cognitive effort and misalignment resulting from each level of fidelity. Lastly, the study synthesizes these insights into pragmatic recommendations for designers.

2. Literature Review

A. Prototyping in Human-Computer Interaction

Low fidelity prototypes make use of quick and simple designs on a medium unlike the final medium such as paper, sticky notes, etc. These can be modified easily and are inexpensive to design. Similarly, medium-fidelity wireframes highlight elements such as layout and structure with minimal focus on visual elements - usually represented using placeholders such as default text, and grayscale boxes and lines. Additionally, they allow for defined workflows such as clickable elements in order to demonstrate user interaction. Wireframes involve a moderate amount of effort and can also be modified easily. During this stage of the development process, feedback from the client and collaboration between stakeholders allows for clarification of the structure of the product before committing to a final design. Lastly, high-fidelity prototypes resemble the final product and involve elements of both visual realism and interactive elements. Generally, these prototypes do not include a database back-end and instead provide a front-end system with partial functionality by simulating data functions. These characteristics of high-fidelity prototypes do not allow for easy modification and hence are expensive to create. Ideally, these prototypes are designed at a stage during the process where the usability and functions of the final system must be approved and validated by the client.

Existing empirical studies in the field such as Virzi et al. [3], who observe usability issues but fail to evaluate cognitive workload and mental-model fit, Walker et al. [6] compare paper sketches and static pages but do not observe specific task workflows involving a greater amount of complexity with multiple steps, and Black, Wilcox, & Minor [7] who also evaluate prototype forms but do not investigate mental models and cognitive workloads.

B. Cognitive Framework

Mental models refer to representations that people may construct in order to comprehend, predict and interact with the world. These mental models as simple representations serve as the premises for problem solving in more complex scenarios [2]. This further highlights the importance of mental-model alignment: when design choices are optimized to users' mental models, usability of interfaces is enhanced. Empirical evidence [6] indicates that probing in Likert questionnaires demonstrated a perceived fit between prototypes and participant expectations. The cognitive load theory is one that describes how working memory may impact learning and problem solving [8]. The NASA-TLX questionnaire is the instrument administered to observe cognitive workloads with scales measuring mental demand, physical demand, temporal demand, performance, effort and frustration. Studies using this questionnaire [9] suggested that higher fidelities were more likely to reduce cognitive workload.

C. Test Case

The scenario of online flight booking employed in this study is of moderate-complexity – with multiple steps that involve decision-making, data entry and memory recall. The usage of multiple steps links to the cognitive load theory due to task complexity and the design and instructions provided to the participants.

Prior research on online booking [10] identified areas of error such as date selection and passenger configuration along with decision criteria that emphasizes trust and clarity of the given options. Despite online flight booking being a well-investigated field, it appears that there is a significant lack of empirical evidence that investigates how prototype fidelity may affect mental-model alignment and perceived cognitive workload in that workflow.

D. Synthesis

The above findings suggest that high-fidelity prototypes lead to decrease in the mental effort exerted by individuals but with resource allocation as trade-off while low-fidelity prototypes are inexpensive to design but jeopardize mental-model alignment. This suggests a lack in empirical research on mental-model alignment and perceived cognitive workload for a moderate-complexity realistic scenario. This study addresses this empirical gap.

3. Methodology

The study utilizes the within-subjects design such that each participant observes each of the three prototypes - the low-fidelity prototype as a paper sketch, the medium-fidelity prototype or wireframe as a clickable interface and the high-fidelity prototype as a front-end static interface. In order to eliminate factors such as order effects, a Latin-square counterbalancing scheme is employed such that participants are administered any of the three orders: paper sketch-wireframe-static page, wireframe-static page-paper sketch, or static page-paper sketch-wireframe.

The participants chosen for the study consisted of 25 undergraduate students between the ages of 18 and 22 with reported previous experience of using online booking platforms. Participants were allocated using opportunity sampling through campus board announcements. This method allowed for ease of recruitment due to quicker allocation of participants. Additionally, demographic information such as participant age, gender, and prior experience with online booking was documented in order to define the characteristics of the sample for this study.

A. Materials

As part of the moderate-complexity task scenario, participants were instructed to book a round-trip flight from Dubai to Ottawa for two adults in business class, departing on July 18th, 2025 and returning on July 27th, 2025.

The low-fidelity prototype as the paper sketch was designed on A4 sized sheets of paper such that each screen in the task workflow was hand-drawn one after the other. The sheets were swapped manually with the assistance of a facilitator based on the participants' interactions with the screens. Participants were required to handwrite during certain steps of the task - done using pencil such that once recorded, the sheets could be erased and reused with the rest of the sample. The medium-fidelity wireframe was designed in Figma using elements such as grayscale visuals and placeholder text with clickable buttons and input text fields. Participants were expected to navigate using the clickable buttons and click on input text fields which were pre-filled and not editable. The high-fidelity wireframe was designed using HTML in order to create a front-end static page (without a database) identical to the final product. Participants were expected to navigate this prototype similar to how one would interact with an actual online flight booking website due to its interface that is similar to the final product. The time taken to create each of these prototypes was recorded: paper sketches were designed within 1 hour, the wireframe was designed within 3 hours, and the static front-end page was designed in 6 hours.

B. Measures and Instruments

Metrics observed in this study include completion time, which was measured in seconds as the time taken for each participant to complete the given task, along with error count, which was measured to be the number of times each participant made a mistake during the task workflow. 'Mistakes' refer to actions such as those that do not align with the given instructions, or more technical ones such as mis-clicks. A NASA-TLX questionnaire is administered to each participant after a prototype has been finished. It contains scales that measure metrics such as mental demand, physical demand, temporal demand, performance, effort, and frustration and the score associated with the perceived cognitive workload is the unweighted average of the above metrics. In addition, a second mental-model alignment questionnaire is administered that asks participants to rate statements on a Likert scale and is filled out after the NASA-TLX questionnaire. The overall alignment rating is the average of the rating associated with each of the statements.

C. Procedure

Prior to the study, participants were provided with a consent form detailing their rights as participants along with a description of the study that they were expected to read and sign accordingly. Participants were then provided with a survey on demographics and previous experience with online booking sites, after which they were debriefed on the aim and procedure of the study with a standardized set of instructions. Each of the participants was then provided with one of the three prototypes based on the prior counterbalancing orders and a specific set of instructions regarding that prototype was read out. Participants were allocated 7 minutes to complete the provided task using the given prototype. Paper sketches involved participants handwriting and circling elements in order to simulate filling text fields and clicking buttons; the wireframe involved participants clicking on text fields and buttons likewise in order to simulate navigation and filling fields while the static front-end page involved participants interacting with real forms containing editable text fields and clickable buttons. Completion time for the prototype and any errors were then recorded. Finally, the NASA-TLX and mental-model alignment questionnaires were then administered to the participants with 5 minutes allocated to each questionnaire for completion. After a 3-minute break provided to reduce the possibility of carryover effects, the other prototypes as separate conditions are administered to the participants in a similar manner. Overall, each condition was allocated a period of 20 minutes, with the overall procedure for each participant taking around 60 minutes or 1 hour.

In order to address the research questions and their hypotheses, the study analyzes the resulting data through initial preparation and cleaning where conditions involving participants failing to complete the task or technical issues are excluded. Secondly, the averages and standard deviations for each fidelity are calculated and outliers (trials that exceeded beyond two units of the standard deviation) are logged and excluded accordingly. For the perceived cognitive workload, each of the participants' unweighted NASA-TLX scores are averaged and a one-way repeated-measures ANOVA with the prototype fidelity is conducted. Similarly, for the mental-model alignment, the average alignment scores are calculated for each level of fidelity and repeated-measures ANOVA is conducted. Lastly, the completion time and error count metrics are summarized for each fidelity. Once the quantitative analysis has concluded, the η^2 for each ANOVA is logged in order to demonstrate the variance by fidelity. Additionally, Cohen's d and 95% confidence intervals are logged in order to measure the degree of differences. Lastly, Pearson's r is computed between the NASA-TLX and mental-model alignment scores in order to evaluate the correlation between the perceived cognitive workload and the mental-model alignment and is further demonstrated using graphs.

4. Results

Participants consisted of 20 undergraduate students between the ages of 18 and 22, with 10 female individuals and 10 male individuals. They reported prior experience of using online booking platforms, after which they completed all three prototype conditions.

A. *Task Performance*

Completion times and error rates were recorded and then summarized as shown in table 1.

Conducting a one-way repeated-measures ANOVA on the completion time metric demonstrated that prototype fidelity had a significant effect with degrees of freedom $F(2, 38) = 1{,}124.5$, $p < .001$, and $\eta^2 = .98$. Following ANOVA, the post-hoc Bonferroni tests demonstrated that medium-fidelity wireframes were more likely to be completed significantly faster than low-fidelity paper sketches and high-fidelity static pages (both with $p < .001$) along with the high-fidelity static page being more likely to be completed significantly faster than low-fidelity paper sketches ($p < .001$). Additionally, error counts were shown to be minimal in nature – none for the low and high-fidelity prototype conditions and although non-zero values for the medium-fidelity condition ($M = 0.45$), they did not diverge significantly from the other conditions, $\chi^2(2) = 3.20$, $p = .20$ (Friedman test).

B. *Perceived Cognitive Workload*

Resulting data demonstrated that NASA-TLX scores decreased with an increase in the level of fidelity.

Table 2
Mean and standard deviation values of alignment scores by prototype fidelity level

Fidelity level	Mean	Standard deviation
Low – paper sketch	3.56	0.46
Medium – wireframe	4.51	0.23
High – static front-end page	4.82	0.17

Conducting a repeated-measures ANOVA demonstrated the significant effect of fidelity on the NASA-TLX scores with degrees of freedom $F(2, 38) = 472.3$, $p < .001$, $\eta^2 = .96$. Further pairwise comparisons demonstrate that medium-fidelity wireframes are more likely to demand a reduced cognitive workload compared to low-fidelity paper sketches ($p < .001$) while high-fidelity static pages are more likely to demand a reduced cognitive workload compared to medium-fidelity wireframes ($p = .02$).

C. *Mental-Model Alignment*

Resulting data from participants' self-reported data demonstrated that alignment increased with an increase in the level of fidelity.

Conducting repeated measures ANOVA demonstrated the significant effect of fidelity on mental-model alignment with degrees of freedom, $F(2, 38) = 286.7$, $p < .001$, $\eta^2 = .94$. Following ANOVA, the post-hoc Bonferroni tests demonstrated that medium-fidelity wireframes were more likely to be aligned closely in comparison to low-fidelity paper sketches ($p < .001$) and high-fidelity static pages were more likely to be aligned closely than medium-fidelity wireframes ($p = .03$).

D. *Perceived Cognitive Workload & Mental-Model Alignment*

Furthermore, conducting a Pearson correlation analysis demonstrated that there exists a moderate negative relationship between the NASA-TLX and mental-model alignment in the high-fidelity prototype condition with $r = -.29$, $p = .04$ that suggests that lower perceived cognitive workload is linked to a better mental-model fit. On the other hand, correlations in the low-fidelity prototype ($r = -.05$, $p = .82$) and medium-fidelity prototype ($r = -.12$, $p = .60$) conditions were non-significant.

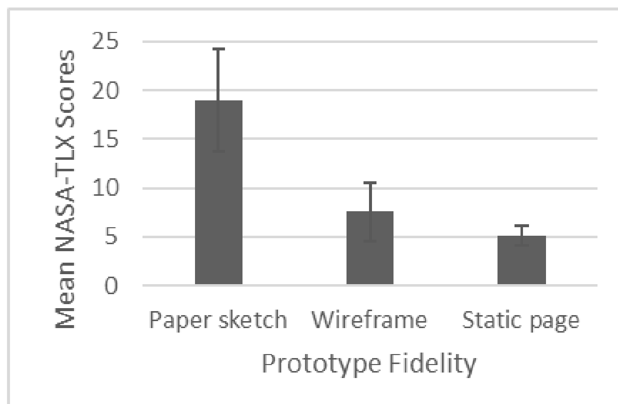


Fig. 1. Mean NASA-TLX Scores by prototype fidelity

5. Discussion

These findings across levels of prototype fidelity therefore suggest a few definite trends. Firstly, an increase in the level of prototype fidelity results in improvements within the perceived cognitive workload and mental-model alignment. Specifically, a decrease in cognitive workload is observed in the repeated-measures ANOVA with the fidelity level significantly impacting the NASA-TLX scores. Low-fidelity paper sketches demanded the highest level of mental effort, while medium-

Table 1
Mean and standard deviation values computed for completion time and error count

Fidelity Level	Mean Completion Time	Standard deviation (completion time)	Mean error count	Standard deviation (error count)
Low – paper sketch	5.85	0.49	0.00	0.00
Medium – wireframe	1.32	0.26	0.45	0.51
High – static front-end page	2.32	0.22	0.00	0.00

fidelity wireframes reduced the amount of mental effort by over 60% and the high-fidelity static pages demanded the least amount of mental effort. Subsequent post-hoc testing established that each decrease in fidelity level led to a significant decrease in workload. Secondly, mental-model alignment scores appeared to increase with a decrease in the amount of workload. It was also evident that fidelity had a substantial impact on mental-model alignment, suggesting that prototypes with higher fidelity were more likely to match participants' mental models, or expectations, of an online flight-booking platform. Finally, observations indicated that participants were more likely to complete the medium-fidelity wireframes in the shortest frame of time compared to the other conditions. Additionally, these prototypes involved a non-zero error count while low and hide-fidelity prototypes appeared to be error free although requiring a longer time frame to complete. These findings hence address the previously stated research questions and validate the hypotheses.

Prior research in the field [1], [6] highlights the idea that the usage of high-fidelity prototypes may optimize usability and user experience but may result in increased expenditure. The findings of this study expand on this idea by presenting evidence regarding the decrease in cognitive workload and improvements in mental-model alignment within a multi-step, moderate complexity, real-life scenario that addresses the empirical gap. RQ1 investigates whether an increase in fidelity causes a decrease in the perceived cognitive workload: the significant effect of fidelity on the NASA-TLX scores and subsequent post-hoc tests support H1 – that higher levels of fidelity do result in lower perceived workloads. RQ2 investigates whether an increase in fidelity enhances mental-model alignment: the effect of fidelity on the mental-model alignment scores and pairwise comparisons support H2 – that higher levels of fidelity do result in higher metal-model alignment scores.

Based on the outcomes of this study, and in order to inform future design practices, the following design measures are recommended. Firstly, the usage of paper sketches may assist in brainstorming and exploring ideas in a limited amount of time with an increase in cognitive workload as a trade-off. Secondly, implement clickable wireframes in order to validate core workflows within the system which subsequently realizes a large part of the benefits associated with usability. Finally, employ high-fidelity static pages specifically for walkthroughs with stakeholders or when verifications regarding the UI are required.

While the findings support the conclusion, certain limitations must be noted. Participants in this study were allocated using convenience or opportunity sampling that may not be generalizable to the wider population, resulting in sampling bias. Moreover, the representative moderate-complexity task of booking flight tickets may be considered too specific, and findings could possibly differ in the context of tasks of varying difficulty. An avenue for further research may involve replication of the study with modified samples such that varied demographics are taken into consideration. Furthermore, longitudinal studies may be conducted in order to document how choices in fidelity in the early stages of development may affect the process and measure product usability over time.

6. Conclusion

Prototyping efficacy is crucial in the early stages of development that transforms user requirements in a conceptual model; however, designers currently lack empirical evidence that may serve as guidelines to determine the appropriate prototype fidelity to balance resource allocation against user mental-model alignment and perceived cognitive workload. The findings of this study demonstrated that an increase in prototype fidelity led to a significant decrease in perceived cognitive workload and improved mental-model alignment. Based on these results, this study recommends an evidence-based approach involving low-fidelity paper sketches in order to explore ideas, medium-fidelity clickable wireframes that verify core task workflows and high-fidelity static pages specifically for stakeholder demonstrations in order to confirm final usability features. In essence, the application of quantitative evidence in the decision-making of prototype fidelity types can allow design teams to optimize their development workflows by detecting usability issues and refining features accordingly.

References

[1] C. Snyder, *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003.

[2] P. N. Johnson-Laird, Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness. Cambridge, UK: Cambridge Univ. Press, 1983.

[3] R. A. Virzi, J. L. Sokolov, and D. Karis, "Usability problem identification using both low- and high-fidelity prototypes," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Vancouver, British Columbia, Canada, 1996, pp. 236–243, doi: 10.1145/238386.238516.

[4] T. Sefelin, P. Giller, and M. Tscheligi, "Paper prototyping—what is it good for? A comparison of paper- and computer-based low-fidelity prototyping," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI '03)*, 2003, pp. 778–779.

[5] R. Molich and J. Nielsen, "Improving a human–computer dialogue," *Commun. ACM*, vol. 33, no. 3, pp. 338–348, 1990.

[6] M. Walker, L. Takayama, and J. A. Landay, "High-fidelity or low-fidelity, paper or computer? Choosing attributes when testing web prototypes," in *Proc. Hum. Factors Ergon. Soc. Annu. Meeting (HFES 2002)*, 2002, pp. 661–665.

[7] A. Black, D. Wilcox, and P. Minor, "Evaluating the effectiveness of rapid prototyping techniques for web applications," *Int. J. Hum.–Comput. Interact.*, vol. 18, no. 3, pp. 231–247, 2005.

[8] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cogn. Sci.*, vol. 12, no. 2, pp. 257–285, 1988.

[9] K. Hornbæk and E. L.-C. Law, "Meta-analysis of text editor usability: How far have we really come?," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI '07)*, 2007, pp. 205–214.

[10] P. Zhang and G. M. von Dran, "Satisfiers and dissatisfiers: A two-facet model of user attitudes on website quality," in *Proc. IEEE Int. Prof. Commun. Conf.*, 2000, pp. 148–155.

Appendix 1: Consent Form

This study is aimed at investigating the effect of decreasing rehearsal time on memory recall. In order to participate in this research study, it is necessary that you give your informed consent. By signing this informed consent form, you are indicating that you understand the nature of this research study and your role in the research and that you agree to participate in the research. Please consider the following points before signing.

As researchers, we ensure the following conditions:

1) We will not put you in danger of physical or psychological harm.
2) We will publish data gained from this investigation anonymously and with respect for your privacy.
3) This data will not be utilized to harm anyone.

Please sign the declaration below:

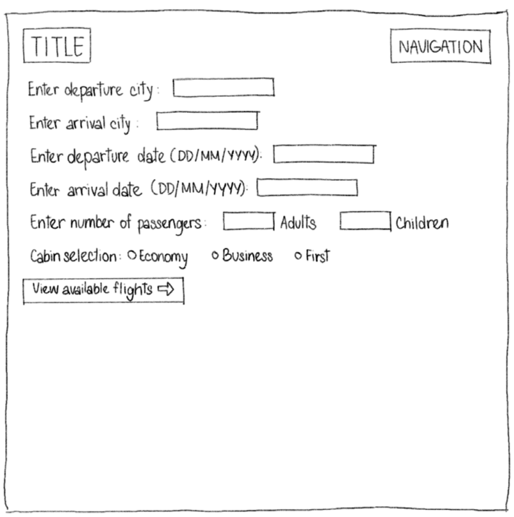
1) I have been informed about the nature of the experiment.
2) I understand that my participation is voluntary.
3) I may withdraw from the study at any time and request that my data not be used in the experimental results.
4) I have the right to a debriefing about the general results of the study, and I may obtain my individual results upon request.
5) I give my consent knowing that all aspects of my participation will remain confidential and that I will not be subjected to any harm or deception.
6) I understand that the experiment has potential benefits.

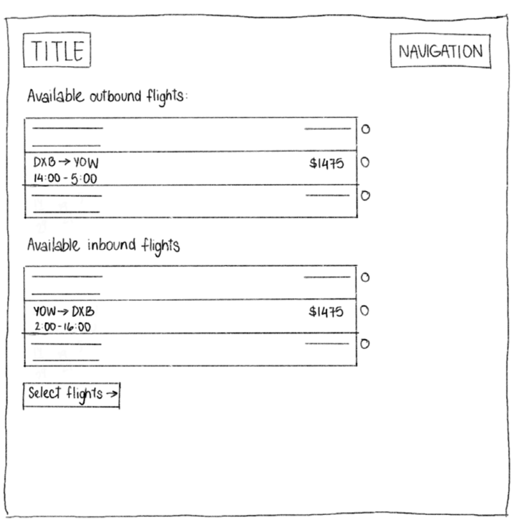
Signature: ______________________________________
Name: ____
Age: ____
Date: ____

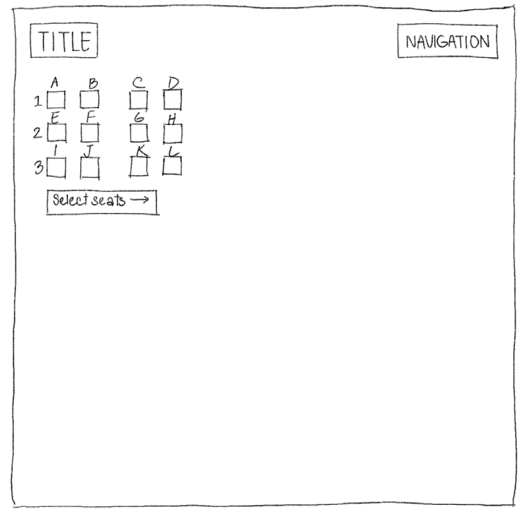
Appendix 2: Pre-Study Survey

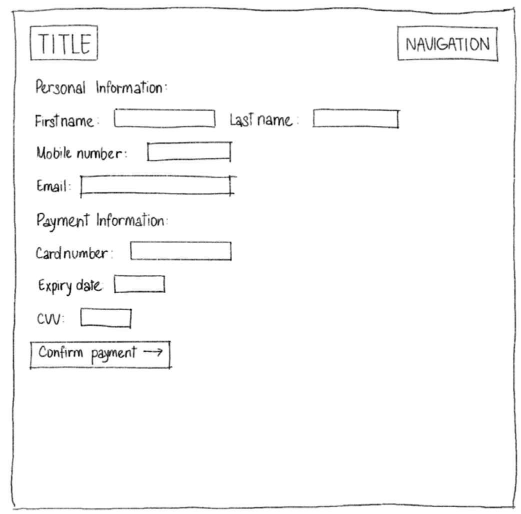
1) What is your age?
2) Select a gender: Male/Female/Prefer not to say/Other:
3) Do you have any prior experience in using online booking platforms: Yes/No

Appendix 3: Low-Fidelity Wireframe (Paper Sketches)









Appendix 4: Medium-Fidelity Wireframe (Figma)

https://www.figma.com/proto/bDMwf90mFe3Zuhusslzwvu/R1---Wireframe?node-id=0-1&t=IJ5rYGWDh03eX6lr-1