

User Study Report **TODO: Set Title (e.g. Using Templates to create efficient calendar apps?)**

Adam Dada, Rafael Fernandes, Rahul Kaundal, Oliver Lehmann, Matthias Lott, Moritz Mani
ETH Zürich

1 INTRODUCTION

As part of the HCI course at ETH Zurich, we went through the human centered design pipeline. From needfinding and ideation to three iterations of prototypes, we found a problem that current users have, brainstormed solutions and in the end came up with two prototype variants that were compared in a user-study.

One main need that users have is that interactions with a calendar should be swift and require little effort. Through several iterations of prototypes, we found this problem can be solved using Templates.

Templates are a special kind of event state that already contain some information such as time, location or participants. They help speed up the event creation process because only some remaining fields need to be filled to create an event, others are already populated. For example, the location of a doctor's appointment will very likely always happen at the same location and can be specified by a template.

Interacting with the templates is very important. Every time a user wants to create an event or modify a template, they are prompted to select a template. Thus, it is essential to get the design of the template selection screen right. For this, we came up with two variants (Figure 1):

- Variant A: This variant comes in the form of a grid view that provides a higher information density. Templates are displayed in two vertical rows of small squares that contain the necessary metadata such as date, location, timing and more.
- Variant B: This variant is designed as a vertical list of full-width template items with their respective metadata. It puts a focus on making the different templates more digestible and the screen less crowded with information.

Other than the interfaces for template management and selection using a different layout, both variants are identical. This was ensured by implementing them both in the same application and adding a toggle to switch between them at will.

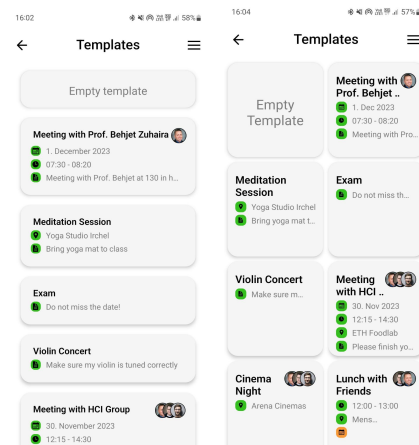


Figure 1: Template Selection in variant A (left) and B (right)

Old:

As it relates to our prototype, the two variants of how to illustrate templates were both incorporated into the same application such that we could programmatically toggle between them based on the user test at hand. The remainder of the app (i.e. everything apart from the template screens) was identical for both variant A and B. The identity also applies to the mechanism of how performance analytics were gathered and exported throughout the study.

2 STUDY DESIGN

The study was designed around the question of how the layout of the template selection screen influences the nature, ease and efficiency of the users' interaction with the prototype. We came to this decision due to the fact that we had imagined the screen layout to play a pivotal role in the efficiency of the event creation process as the number of stored templates increased. Furthermore, since the screen for selecting templates is used for template editing as well as creating, finding the correct layout would yield significant reward in terms of speed and **mental load** when performing certain user actions.

2.1 STUDY SETTING

There was no specific hardware used, all studies were conducted on the personal devices of the group members (running either iOS or Android). Similarly, there was no consistent environment for the participants. **Many studies were conducted at home, some of them were done remotely.**

2.2 VARIABLES AND HYPOTHESES

The only independent variable is categorical and encodes the interface variant (A or B). **We made sure that memorization ability is not an independent variable by shuffling the list of templates every time the user needs to select one.**

The study analyzes four dependent variables:

1. *Actual Speed*. This indicates the time that a user spent on the template selection screen. It is an objective, interval variable.
2. *Perceived Speed*. This indicates the time a user thinks they spent on the template selection screen. It is the subjective counterpart to actual speed and also an interval variable.
3. *Scroll Amount*. This indicates how far the user scrolled while looking for the correct templates. It is an objective, interval variable.
4. Usability. This indicates how usable a system is and is captured using 10 questions (SUS questionnaire, **TODO: Link**), resulting in an overall usability score between 0 and 100. The 10 questions are subjective, ordinal variables. The final score is **TODO**.
TODO: Rephrase. Quote something?

These four dependent variables naturally induce the following four null hypotheses:

1. There will be no effect of changing the grid view to a list view on the time a user needs to select a template (actual speed).
2. There will be no effect of changing the grid view to a list view on the time a user thinks they need to select a template (perceived speed).
3. There will be no effect of changing the grid view to a list view on the amount that the user has to scroll (scroll amount).
4. There will be no effect of changing the grid view to a list view on the usability of the template selection screen (usability).

2.3 PROCEDURE

Participants were recruited using a method called convenience sampling (https://en.wikipedia.org/wiki/Convenience_sampling). Information about the participants can be found in section 2.4.

To test whether the null hypotheses hold, a task consisting of two subtasks was devised:

1. Edit the “Lunch with Friends” template by setting its color to green.
2. Plan a trip to the zoo using the template “Zoo” next saturday lasting the whole day.

In order to compare the two design variants and find out if the hypothesis can be rejected, a counterbalanced within-subject study was conducted. This means that every participant tests both variants, half of them starting on variant A, the other half on variant B. The study consists of the following steps, further details can be found in the Study Protocol (https://github.com/eth-hci-course/hci-project-hci2023-group-12/blob/2023/Deliverables/Study/Study_Protocol.pdf):

The user starts by filling out a questionnaire in which *participant data* (age, gender, technical affinity and familiarity with digital calendars) is collected anonymously. Then, they are presented with the task and asked to practice it once on both variants.

Afterwards, they perform the task on the first variant during which we collect *performance data* (time spent selecting a template and vertical scroll amount). Subsequently, the user fills out a questionnaire where *subjective ratings* (perceived time spent selecting a template and usability) are collected. This is then repeated for the second variant. During the whole study, *qualitative data* (feedback from the user, what the user struggles with, ...) is gathered and discussed with the participant in the end.

Old:

Question: Should the information about study setting (used hardware and environment) belong here?

To conduct the user study we had four different types of collectable data at our disposal:

1. *Participant data: We collected basic demographic data (age, gender), technical affinity and the user's familiarity with digital calendars.*
2. *Subjective ratings: We prepared questionnaires containing questions to register the mental load, the usability of the design, and the perceived speed when using the prototype.*
3. *Performance data: We measured the following metrics on the template selection screen: number of seconds spent on the screen, how much users scrolled vertically on the screen in terms of pixels, and number of times users clicked on the wrong template.*
4. *Qualitative feedback: We asked users for any kind of feedback that came to their minds. In addition to that, we questioned them about particular things that occurred during the test and we wanted to know more about.*

Question here: do we also need to write about the tasks that we had users perform?

2.4 PARTICIPANT DISTRIBUTION

For this study, 12 participants were interviewed. They have an average age of 44.333 years (SD = 21.073 ys). 7/12 participants identified themselves as female, 5/12 as male, 0/12 as other/none.

They rate their technical affinity with a median of 6.

4 people never used a calendar app before, 3 rarely use one, 4 people use them often, 1 person uses a calendar app daily.

3 RESULTS

As will be explained in a later section, the results that were obtained were heavily impacted by a suboptimal participant distribution. We noticed a trend that younger testers tend to favor variant A more strongly, which might have led to statistically significant results.

Question: Also mention the effect size?

TODO: Pictures

3.1 ACTUAL SPEED

To compare the effect of the **interface variant** on the **time spent on the template selection screen**, we conducted a **paired samples t-test** (all Shapiro-Wilk $p > 0.05$ and Levene's $p > 0.05$).

With interface variant A, participants on average needed 40.837 seconds (SD = 12.935 sec) to select the templates. With interface variant B, participants on average needed 46.132 seconds (SD = 16.365 sec) to select the templates.

The mean difference between the two groups was **not** statistically significant; $t(11) = -1.384$, $p = 0.194$.

3.2 PERCEIVED SPEED

To compare the effect of the **interface variant** on the **time participants think they spent on the template selection screen**,

we conducted a **paired samples t-test** (all Shapiro-Wilk $p > 0.05$ and Levene's $p > 0.05$).

With interface variant A, participants thought that they on average needed 6.167 seconds (SD = 3.486 sec) to select the templates. With interface variant B, participants thought that they on average needed 7.167 seconds (SD = 4.174 sec) to select the templates.

The mean difference between the two groups was **not** statistically significant; $t(11) = -1.864, p = 0.089$.

3.3 SCROLL AMOUNT

To compare the effect of the **interface variant** on the **amount that the user scrolled on the template selection screen**, we conducted a **wilcoxon signed-rank test** (Shapiro-Wilk (Variant A) $p > 0.05$, Shapiro-Wilk (Variant B) $p < 0.05$ and Levene's $p > 0.05$).

With interface variant A, participants on average scrolled 1448.417 pixels (SD = 1064.636 px) while selecting the templates. With interface variant B, participants on average scrolled 2116.583 pixels (SD = 2117.855 px) while selecting the templates.

The mean difference between the two groups was **not** statistically significant; $Z = -0.863, p = 0.388$.

3.4 USABILITY

To compare the effect of the **interface variant** on the **usability (Total SUS score)**, we conducted a **wilcoxon signed-rank test** (ordinal data).

With interface variant A, participants on average rated the usability of the template system with 85.833 points (SD = 13.831). With interface variant B, participants on average rated 83.333 points (SD = 12.123).

The mean difference between the two groups was **not** statistically significant; $Z = -1.358, p = 0.174$.

3.4 QUALITATIVE FEEDBACK

A variety of qualitative feedback has been recorded as part of the user study and the majority of it relates to, firstly, the Template Selection and Overview screens, and, secondly, the New Entry screen.

Concerning the former, testers have been vocal about the fact that with a rising number of templates, a mechanism to sort templates would be advantageous and probably lead to a shorter search for a specific template. This tendency to process information in an efficient way was also apparent in the participants' remarks about liking the grid version of the templates view better in comparison to the list view. **Using the grid layout, users had the feeling of having access to more data at a glance and perceiving a shorter time duration to locate specific templates.** However, it was also reported that even though the screen held more information, the zig-zag mechanism of sifting through the many entries was not as pleasant as on the list version of the templates view. There, the flow of processing information goes directly from top to bottom

and, as a result, decreases the chances of missing particular templates which would cause one to scroll back to already seen information.

Furthermore, some users seemed to be confused whether changes made when editing templates were actually permanent. **A snackbar or another form of visual input of editing success would have provided reassurance about the actual implementation of changes.** Interestingly, also the grid or list item on the template selection and overview screen which was supposed to indicate the choice of a blank template or event entry caused a certain confusion among user study participants. **Hence, it took users some time to perform the correct action to create a new template or event entry when prompted.** Apart from the mentioned shortcomings, however, users were content with the template views and regarded them as easy to use.

As it relates to the latter, the verbal feedback of the user study participants agreed on the fact that the green checkmark on the right side of the top app bar was falsely implying auto-saving functionality. Instead of representing an icon button to save any changes which were made to the current entry or template, many users believed the green checkmark to be confirmation that their changes were registered and saved. In retrospect, a text button with an instructive label such as 'Done' or 'Save changes' would, according to testers, have resulted in a clearer picture of what to do once changes were made to templates or new entries.

4 DISCUSSION

4.1 ACTUAL SPEED

There is no indication that template selection on interface variant A is faster than on variant B. H_0 is not rejected. A reasoning for this might be that the two variants have different factors that hinder user speed. Variant A had participants scan the templates in a zig-zag fashion instead of a straight line. Variant B has templates displayed less densely, requiring more eye movement over all. **This is supported by a user remarking that they were slowed down by the zig-zag reading.**

Actual Speed:

- Big Variance due to old and young ppl (high SD). Missing ppl between 30 -50 yo.
- Lags
- Grid: Reading was hard, List: Less dense information, more eye movement required

4.2 PERCEIVED SPEED

There is no indication that users perceive template selection as faster on interface variant A than on interface B. H_0 is not rejected. The two variants may be perceived similarly, because users are performing the same actions in a similar intensity (scrolling, reading and tapping).

However, because we observed a p-Value of 0.089 which is quite close to 0.05, we also provide a different interpretation:

In case a user never has to scroll, there are indeed less actions to complete (no scrolling). As variant A has a more dense visualization (four times no scrolling was needed, in variant B this only happened twice) the perceived speed might be slightly lower.

As a note, *user-perceived speed* is lower than *actual speed* because firstly, testers probably did not consider app response-time as part of the perceived speed and secondly, testers most likely didn't include the time it took them to navigate from the detailed template summary to event creation/template modification which was recorded in the *actual speed* metric.

Perceived Speed:

- Best results: 8% p-Value (5% wanted). Grid was faster.
- People probably didn't count the lag and didn't count the time spent on the detailed template view
- Lags
- Denser View: Probability that template was already on screen is higher

4.3 SCROLL AMOUNT

There is no indication that users scroll less on interface variant A than on variant B. H0 is not rejected. A reason for this might be that users feel less secure on variant A and start asking themselves if they missed the template and scroll back to already visited sections to recheck whereas in variant B this is not the case. A participant remarked that variant B is calmer and the reading flow goes cleanly from top to bottom.

Scroll Amount:

- Grid View needs more scrolling than expected because users make mistakes or feel like they did and recheck..
- Users did not comment on their scroll behavior. depended highly on the shuffling.
- Lags. Target Template moved to a section they already scanned after the shuffle
- Denser View: Probability that template was already on screen is higher

4.4 USABILITY

There is no indication that interface variant A is more usable than variant B. H0 is not rejected. As can be seen in Figure X (TODO: Likert Scale), users rated the usability of the interface variants almost identically, this might be because there simply is not much difference in the interface variants. After all, users still perform the same actions (scrolling, reading and tapping). Some minor differences in favor of variant A can be seen in Frequency and Integration. It is worth stating that a user mentioned they like that they can see more information at a glance in variant A.

Usability:

- Users did not perceive much difference in usability.
- Frequency and Integration had some "bigger" changes in favor of Grid View.

TODO (Interpret/explain the results, Maybe also refer to related work)

5 LIMITATIONS

Reasons for good results with young ppl:

- Young users are not hindered by zig-zag reading

- Young users can more quickly scan and select the template (if they need to scroll, they are inconvenienced).
- Young users need to scroll less, they don't make scrolling mistakes

No real build (lags, not using your own device)

Small sample size (and weird distribution)

Different setting (zoom, at home, operating system, device height, using a foreign device, time of day)

(Presenter/User Error)

Not enough templates used to really measure a difference

TODO (Note all the limitations, why some things didn't work, many additional variables, ...)

6 FUTURE WORK

Either: Redo study? Prescreener to only take frequent app-users? More participants?

Or: Directly stick with variant A, implement the various feedback that was given (floppy disk icon, ...)

Additionally: Improve the study. Ask for users for scan and scroll behaviour? Conduct study for different amounts of templates? Make them not exercise all tasks every time

TODO (What would you want to try in future work. What would you want to fix.)

7 CONCLUSION

Variant A has higher load but that's not a problem because users can most of the time handle it and be faster?

Younger Generation has more focus on the task (because they are familiar with tech), Older generation has less focus on the task (because they need to focus on using the device) and have less mental capacity for the grid view.

TODO (Conclusion and take home messages for readers and yourself)

REFERENCES TODO

- [1] J. Allebach. Binary display of images when spot size exceeds step size. *Applied Optics*, 15:2513–2519, August 1980.
- [2] E. Catmull. A tutorial on compensation tables. In *Computer Graphics*, volume 13, pages 1–7. ACM SIGGRAPH, 1979.
- [3] Peter Litwinowicz and Lance Williams. Animating images with drawings. In Andrew Glassner, editor, *Proceedings of SIGGRAPH '94* (Orlando, Florida, July 24–29, 1994), Computer Graphics Proceedings, Annual Conference Series, pages 409–412. ACM SIGGRAPH, ACM Press, July 1994.

Qualitative feedback notes (for Adam):

templates screen:

- Did not notice a difference between the two templates..., Keyword- and alphabetical sorting are definitely necessary when we have many templates,
- No alphabetical sorting, should not throw you back to template screen but to landing page, whole edit screen should fit on one page
- Liked Variant A more because it gave a better overview.

- The overall system is not bad, but having many templates definitely requires sorting/filtering."
- icons on template screen are confusing at first because user thinks that he can edit these values. Is this even necessary information to display?
- The color should be displayed in the template screen to check which color is set
- "Problems: Tried to edit directly in the template overview, Symbols in NewEvent look like buttons, color not given in template overview (very confusing), confused edit template with use template
- Debriefing: They found it easy to use, but were confused because the checkmark seemed that changes are already registered (does not look like confirm). They found it fast, but wished for a clearer differentiation between edit and use."
- the refreshing of template overview is confusing, symbols look like buttons, no indication of what changed after editing a template
- "Problems: Empty template is too plain or it is not apparent to represent a new event (took some time finding a way to create a new event)
- Debriefing: They like the Idea of templates, but were confused with empty template. They wish for a way to check the work after editing a template."
- **Extract from this that zig-zag might take a bit longer (actual speed)**
- The user likes the grid view more than the list view, because it can see more data in one glance and the target template is faster to find.
 - **Extract from this that the chances of only having to click are higher (usability & perceived speed)**
- List is better because text is more readable and 'calmer' because the flow goes neatly from top to bottom
 - **Extract from this that a user doesn't have to scroll back and recheck already scanned sections (scroll amount)**

NewEntry Screen:

- "Didn't realise they could scroll the NewEntry screen at first.
- Didn't realise that All Day - All Day was for selecting time if All Day was not set.
- time selection is not registering when not pressing on ok in NewEntry,
- was looking for tick on bottom instead on top
- "Didn't understand that the checkmark stands for save, would prefer a floppy disk icon
- The user would like to use ""Save"" button rather than a tick to save the template (that is, because the user is a iPhone user and there is a ""Done"" button). "
- The tester was a bit confused after both subtasks concerning whether the changes were actually applied to the template or the new event was actually created. The tester would have liked a visual confirmation as to the success of his actions. I guess a snackbar or success screen would have done the job.

General things:

- Open Events are not clear because ""Open Events"" can be interpreted as ""Open a New Event"".
-
- The user would rather use the ""+"" button rather than open the drawer and select template.
- green checkmark is not intuitive because it is on the top. The control flow goes from top to bottom, but at the end you have to go all the way to the top again**Which screen?**
- There should be pictures to uniquely identify events, Sorting has to be implemented **What is meant here?**
- if 2 and 1 was switched they would maybe click the plus."
What is meant here?

Additional (Matthias) (Please include because i reference it in the Discussion):

- Must search titles in a zig-zag pattern which is not nice for the brain