

Stage 4 – Prototype Evaluation

Design a cloud-supported desktop task management interface for post-secondary students to help them organize, view, and plan their tasks.



**Kamyar
Ardekani**



**Francisco
Escalona**



**Antoine
Ponsard**

AI – Evaluation Report

Goals, approach and rationale. The evaluation at Stage 4 was intended to address the key challenges of our interface, as identified at Stage 3. To this end, we conducted a usability study on our medium-fidelity prototype with 12 potential users, to evaluate several key elements of our design approach.

Execution details. We conducted our evaluation with 12 participants (8 Masters, 3 PhD, 1 undergrad) who had not previously seen our prototype. Given the restrictions we defined in our requirements at Stage 2, we selected only participants who were already using some sort of an electronic calendar to manage their tasks. To benefit from triangulation, we used a combination of three methods during our evaluation. The first one was to follow talk-aloud protocol while performing tasks on the med-fid prototype. We started with a simple set of tasks, which did not require prior knowledge of the interface; then we briefly explained the concept of the Main Todo List and Contextual List (CL), before proceeding with a set of more complex tasks. The second part of the evaluation was a short anonymous questionnaire, in which the user had to evaluate the interface on several common criteria. The third and last part was a post-hoc semi-structured interview intended to gather more in-depth feedback from the users, and possibly some ideas for improvements.

Divergence. Because we knew most of our participants, we were expecting them to be biased when answering evaluative questions about our prototype. To reduce this bias, we decided to ask these questions via an anonymous questionnaire, which was not part of our evaluation proposal at Stage 3. We were also planning to use about 5 participants; but, based on the feedback we received from the design review, we decided to increase this number to 12. It gave us a richer picture of our target population, and made our quantitative evaluations (such as the aforementioned questionnaire) more meaningful.

Summary of data. For the talk-aloud protocol, each sub-task performed by the participants was assigned a score from 1 (when the user couldn't perform the subtask) to 4 (when it was performed without any problems). Figure 1 is a heatmap of the distributions of scores for each subtask; the instructions given to the participants for each subtask can be found in Appendix A1.

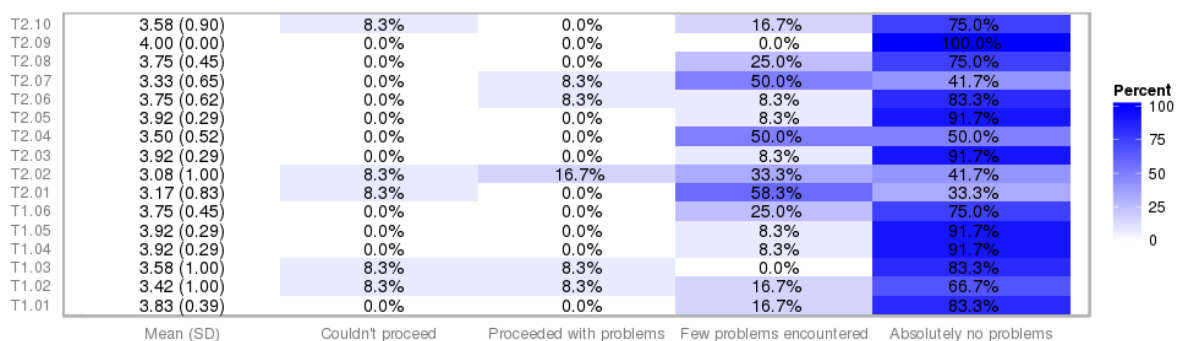


Figure 1. Subtask statistics

Pinning tasks to the Main List (T1.04) was intuitive, which is of particular interest because it is one of the things that we hypothesize as helping students build a better mental model of the connection between different parts of the interface. Similarly, this evaluation showed that our design conveys well the crossing-out of items (T2.05), clicking on the trashbin to clean a list of items (T2.06), and to a lesser degree dragging and dropping reminders in the calendar (T2.10).

As it is, the overall navigation between views (T1.02, T1.05, T2.04) requires a little familiarity with the interface, but it quickly becomes intuitive. Some users expected items on the calendar to link them to the todo lists. We observed some problems using drag-and-drop to reorganize tasks (T1.06, T2.07), where about half of the participants tried to drag an item by clicking on its text description - a behavior that we explicitly prevented when building the prototype, on the ground that it would lead to errors. Therefore it seems that the empty area of an item doesn't provide visible enough affordances for dragging, even if the mouse cursor is changed into a hand.

The most severe problem we discovered is the lack of affordance for scrolling in the calendar (T2.01, T2.02) which was also not visible enough in part due to alternative mechanisms to change the dates, possibly related to transfer effects from other calendar applications.

The questionnaire responses are summarized on Figure 2, and the means and standard deviations for each criterion are shown in Table 1. The reaction seems to be mostly positive, with Usability scoring lower than the other criteria - we will discuss this finding in depth later in the report. It is worth mentioning here that the post-hoc interviews correlate well with the results of this questionnaire.

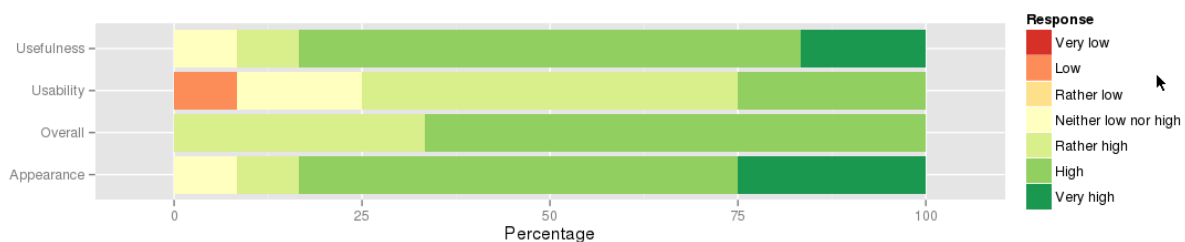


Figure 2. Questionnaire responses distribution

	Usefulness	Usability	Appearance	Overall
Mean (SD)	5.92(0.79)	4.88 (1.13)	6.04 (0.86)	5.71 (0.45)

Table 1. Questionnaire statistics

From the interviews, we learned that some users didn't feel the two main views to be as connected as we wanted them to be; which is a serious problem since we presume this aspect of the mental model would help students to get a better overview of their activities. On a more positive note, the behavior of the Contextual List, a central piece of our design, seems to be adequately inferred from the interface, because students were able to explain its purpose well. For instance, Participant #6 said it was like "[...] having a list of prioritized items without having to write one." With one exception, all participants thought the CL was a useful tool to have. Most people liked the idea of combining calendar and todo lists, and several said they would use such a system to manage their tasks, as long as it was integrated with their current tool - namely Google Calendar or iCal.

Although users liked the look and feel of the todo lists, and the overall appearance of the application, some mentioned that they would prefer to be able to personalize it, for example by changing the background, or the colors of the lists. Other recommendations include the use of a datepicker to input times, in addition to the intelligent text box parsing a date written in natural language (e.g. "next Friday at 5pm").

Conclusions. The idea to have an electronic calendar combined with todo lists was well received, even if the sense of "connectedness" of the two views has still to be improved. Most participants understood the concepts of the Main Todos and the Contextual List, which were both considered to be useful. Therefore, we believe our high-level design approach has been validated at this stage. Students also liked having deadlines in the calendar, especially for academic work, but few understood clearly the distinction we intended between events, deadlines and simple todos - based on whether they have a date, a duration, or no time information at all. These concepts seemed difficult to infer from the current prototype, but were understood easily when explained to the participants. As mentioned before, several adjustments to the interface appeared to be necessary during the evaluation, such as the lack of scrolling affordance of the calendar. These modifications are lesser in scope and can be addressed individually, which would be appropriate to do as the project approaches the high-fidelity prototyping phase.

The evaluation itself had some flaws. First, some of the subtasks presented to the participants were confusing, which interfered with the assessment of the interface. We ran a pilot before the actual evaluation, but some of these issues escaped our attention. Second, there is a problem of ecological validity in our study, since the prototype was pre-populated with arbitrary tasks and events. A better study (perhaps a longitudinal one) would allow students to be familiar with the content displayed by the system. Third, most of our participants were Computer Science students, which introduced an obvious bias in our population, due to the familiarity of these students with complex software. We weren't able to address this last issue further because of time constraints.

BI – Final design rationale and discussion of the state of our design

To present the final conclusions of our project, and the corresponding design recommendations, this section is structured with respect to different elements of the interface.

As mentioned before, the concept and purpose of **the Contextual List** were generally well understood. However, some students couldn't tell precisely how it works, or what type of activities it contains. The CL was regarded as useful for filtering out the least relevant information. Therefore, despite its minor usability issues, we believe it has been validated. Introducing a new concept like this one is inherently difficult for first time users. We believe a simple popup-based tutorial of the interface could be designed for new users, and the differences between basic todo items, deadlines and events could be made clearer throughout the design.

The Main Todo List was easy to find and use. Participants understood and welcomed this idea, thus we think this component has been validated as is.

Integration of calendar and todo lists. Switching back and forth between the calendar and todo lists view was a fast and intuitive task for our participants, and required little cognitive effort once learned. However, some users stated that these two parts of the interface did not feel connected adequately. We have found three possible explanations for this problem. First, the calendar is not just pushed to the left during the transition between the two views (as in a classic film strip), but rather transformed into the agenda, which might be confusing for some users. Second, the prototype was populated with arbitrary items, whereas in the real world users would have created and categorized the items themselves - a process during which they could have noticed how the different parts of the interface are synchronized. Third, not all functionalities that we had planned in this regard were implemented in the prototype; most notably, hyperlinks from calendar to todo lists were not implemented. Overall, this approach shows promise, but still needs some important improvements; one of which is to refine - or redefine - the use of the agenda in the todo lists view, as a representative of the calendar.

Other recommendations. Giving students a more refined approach to select the time span shown in the calendar view and the CL might make the interface more usable. Our recommendation is to prototype the Weekly View on the next iteration of the project, and later explore the idea of a "time lense" that allows students to dynamically focus on and expand a period of time in the calendar.

B 2 – Project-scale Contributions

To understand how students use available tools to manage their time, we conducted an initial exploratory evaluation. From this, two key insights emerged that guided our design of a new solution: no single tool meets all the needs of a student, and there seemed to be a category of students who all use an electronic calendar with some other independent record of tasks or todos. They also considered having a good overview of their activities was important, but not particularly well supported by the current interfaces.

To address these issues, we analyzed the different types of activities students have in their schedule, and created several techniques to represent each of them in an appropriate way. Most notably, the Contextual List displays a prioritized list of a limited number of time-sensitive activities based on the time period the user is focusing on, and the Main Todo List allows the students to pin their more immediate or important todo items to a list that is visible all the time. The former can be seen as an automatic prioritization tool based on the known time management concept called “The Important/Urgent Matrix” (Convey, 1994), while the latter is a more manual prioritization method that accounts for personal differences in task management.

These two new components are combined in a summary column displayed between a calendar and a todo lists view, bridging the gap between these two views in a useful way, as validated in a subsequent evaluation. The overall approach was well received, and several students expressed their interest in having access to such technology.

B 3 – Reflection on our design process

Since the instructions stated that this part could be free-form, we decided to write it as a list of remarks on the different stages of our design process.

- We learned how to conduct effective **brainstormings**, and came up with the following three rules:
 - Always begin with an individual brainstorming
 - When a decision must be made (e.g., between different topics), write all the possible choices on a board, then ask each person to “vote” for items by putting a colored dot next to them.
 - Limit the duration of the brainstorming to two hours, or have long breaks
- Our **task examples** changed a lot during the project, from a high-level, character-based version at the beginning, to a step-by-step, less comprehensive version at the end, which was necessary to conduct the evaluations. But this wasn’t a problem, because we always knew *what* we were trying to support, even if the nature of our interface made it hard to write down in a small set of tasks.
- Preparing, conducting and analyzing the **evaluations** at stage 2 and 4 was very valuable. It showed us unequivocally the value of user-centered design. Choosing

which evaluation methods to use was quick and easy, but we believe our evaluations went well because we iterated a lot on the evaluation instruments themselves.

- We had a lot of choices to make during the process, and it was sometimes hard to leave a promising idea behind.
- Prototype
 - Sketching on paper or whiteboard was extremely useful at the beginning. Then we used **Balsamiq** to do incremental changes without having to re-sketch everything, and also to try the layout of the interface - especially because we had so many constraints on clutter and screen real estate. However, we didn't like Balsamiq that much; it seems that a lot of things are still to be improved (e.g., the sharing features, or the lack thereof).
 - Then, we found no value in making the **low-fidelity** prototype interactive with small pieces of paper. Indeed, most of the controls we used were intuitive enough for any student to imagine them from the Balsamiq mockup.
 - Besides, we found very little value in the **Cognitive Walkthrough** itself. It is too formal and would require much more time to be done properly; whereas we clearly lacked time in Stage 3: only three weeks to come up with a design, create a low-fidelity prototype, then a medium-fidelity one! And if the CW had gone wrong, we simply wouldn't have had time to start again with a brand new design and a new low-fidelity prototype. Therefore, we believe it would be more valuable to replace the CW with design discussions at the very beginning of Stage 3, when things are not set in stone yet. The idea would be to use the first hour of a lab to brainstorm in teams, then rotate people around during the second hour to discuss other groups' ideas. This would be done for both the first design ideas (only rough sketches) and the first draft of the low-fidelity prototype.
 - Our **medium-fidelity** prototype was very complicated to implement, for several reasons. First, it was very broad in scope. Second, the interface was complex, with lots of animations and reactive behaviors. And third, the functionalities we chose to implement required an underlying object representation of todos, which made it more difficult than a regular HTML page. Making the prototype look good was fast and easy at first, but then it slowed down the rest of the process, because you cannot add an ugly element to an already good-looking prototype.
- It is frustrating that we won't incorporate our Stage 4 findings in an improved version of the prototype, to "keep something" at the end of the class. In reaction, it made us put too much effort in the medium-fidelity prototype than what was strictly necessary for testing it, because we knew it would be the last thing we would do in the course of this project.

Appendix A.I – Evaluation Instruments

Evaluation Script – For the evaluator

Setup: not fullscreen 5 rows and enough width

Greetings
Consent form

Participant's information

- undergrad / master / PhD
- Do they use: electronic calendar / electronic todo lists / task management app / other
- Attended the demo?

test the prototype of an interface for student task management
Follow a talk-aloud protocol -> speak what you are thinking

"You will perform two tasks. The purpose of the first task is to discover the interface, so I won't give you any information upfront. After that, I will tell you a few things about the interface, and then you will do the 2nd task."

Talk aloud - part 1 (5 min)

Intervention (2-3 min)

- do you have questions about the interface?
- Explain main todo: "This is what we call the main todo list. You can pin item on it to keep them always visible."
- Explain contextual list: "This is what we call the contextual list. It shows the five most relevant deadlines of the time period currently displayed in the calendar. If you want to see more items, you can expand it by dragging the bottom."

Talk aloud - part 2 (5 min)

"Eventually the system would recognize a date written here, but this is not implemented yet. So I will just do a little magic... and voilà! You can go on."

Mini-questionnaire (7-point Likert scale)

Tell them that we will do a short (5, maybe 10 minutes) interview

Semi-structured post-interview (5-10 min)

- Why did you encounter these problems [noted before] while performing the tasks
- What do you think is the purpose of the contextual list in the application?
- Do you think the contextual list might be useful?
- Can you tell one good and one bad aspect of the interface, if any?
- Would you or wouldn't you use this system as a student? Why?

Evaluation Handout – for the subject

Medium-fidelity prototype evaluation

Task 1

1. Find the exact deadline for the biology essay
2. Find what you have to do for it (in your “Courses” todo list)
3. Find the “read paper on bioinformatics” item
4. Pin this item
5. Go back to calendar view
6. Put this “read paper” item on top of your main todo list

Task 2

1. Check your most important deadlines for November
2. Check all your upcoming deadlines for November
3. Notice the meeting taking place on Monday, November 18th, for your Research Assistantship
4. Review the TODOs related to this meeting (in your “Research” todo list)
5. Cross out “redesign lab webpage” and “review python regex reference”
6. Clean your “Research” todo list
7. Create a new task before “write data parser”: “read the data format”
8. Assign a deadline for this new task on Sunday, November 17
9. Hover on this item in the calendar
10. Set a reminder by drag-and-dropping the reminder icon to the day before

Thanks for participating!

Evaluation Questionnaire – for the subject

Questionnaire

We would like to know your opinion on the interface you've just tested. For each of the following criteria, please circle one number.

Usability

I found the interface:

Hard to use 1 2 3 4 5 6 7 **Easy to use**

Appearance

I think the interface looks:

Bad 1 2 3 4 5 6 7 **Good**

Usefulness

For the purpose of student task management, I believe this interface is:

Not useful 1 2 3 4 5 6 7 **Useful**

Overall opinion

Negative 1 2 3 4 5 6 7 **Positive**

Appendix C.I - Cumulative Contribution Report (CCR)

344/544: CUMULATIVE COMMITMENT ARCHIVE

Version: 1.2

TEAM:

Nursery Web

define week 1 : 3/Sep/13

SECTION:

T1B

Date last updated:

22-Nov

SUMMARY:

Ardekani	Escalona	Ponsard	<Lastname 4>
----------	----------	---------	--------------

Team Role (s):			Project leader	
Cumulative Hours:	396	113	136	147
Percentage:	100%	29%	34%	37%
				0
				0%

Team Comments (optional):

Phase 1:				
Phase 2:				
Phase 3:				
Phase 4:				
Phase 5:				