Visualising the life cycle of stars

Report on second user evaluation

Nikki Rademaker, s2641887
Levi Peeters, s2011174
Assistant: Lisa Pothoven

In the week of December 13th, we held the second user evaluation for our project. For this project a program is being created that visualises the life cycle of stars. Two evaluation groups were used for this user evaluation. The first group consisted of people without an astronomy background, and the second group were astronomy students of different years at different universities. We were able to hold the first user evaluation during a problem class for the course "Stars", but since the classes of this course were finished already, the second user evaluation was split into two groups. This enabled us to measure the usability attributes in person, as well as getting input and feedback from our main target users (astronomy students) online. The usability specification is specified in the design document of this project. An overview of this, and the measured values are shown in Figures 12, 8, and 9. In total 17 evaluators were reached, where 11 were from the first evaluation group, and 6 were from the second evaluation group. Again, for this user evaluation, two empirical methods were used. There were two surveys created to evaluate the program, one for each of the evaluation groups. Additionally, the first evaluation group was observed while executing certain given tasks.

1 Prototype & Instructions

1.1 Prototype

During this user evaluation, all of the main features and interactions of the prototype were implemented. The interface allowed the user to start, stop, and pause the animation of the first few stages of the life cycle of a star. This animation included a working timeline that had corresponding colours to the layers of the star that get formed during each stage (Figure 1). Additionally, the users were able to navigate to a specific stage of the star by interacting with the timeline on the bottom of the screen. The users could also switch between the view of the internal, and external structure of the star (Figure 2). Moreover, the evaluators could view different stars with different masses (Figure 3). For these stars the same interactions were possible.
In the first user evaluation, the program received some feedback to make the interface more clear, and educational. As suggested, by hovering over the objects of the interface with a computer mouse, the users were able to get more information on the functions of these objects (Figure 4). The users were also able to hover over the layers of the star to view which fusion reaction took place in each layer (Figure 5). To get an indication of how much a star grows during its life cycle, references (such as Jupiter and the Sun) were added to the interface (as seen in Figures 1, 2, 3, 4, and 5). In addition, the animation was made smoother to make it clear for the users that the animation was running. Unfortunately, even though it was requested, showing the surface temperature of the star is still in progress. Therefore, it was not part of the interface yet.
At the time of the second user evaluation, not every feature of the program worked optimally yet, but the main interactions were there.

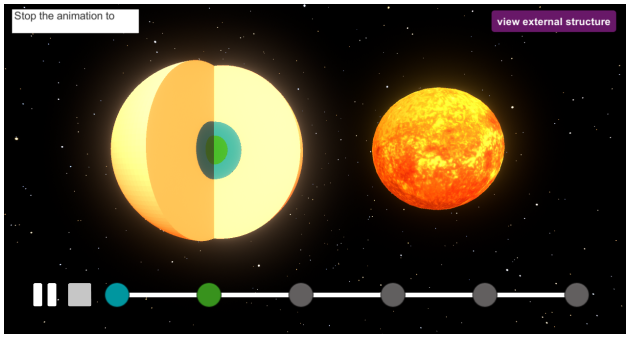

Figure 1: The interface during the animation of the life cycle of a star.

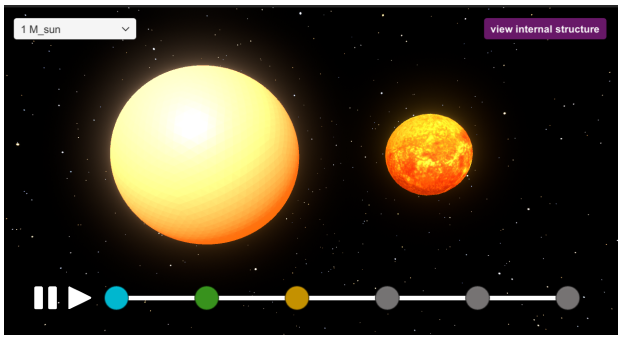

Figure 2: The interface when looking at the external view of a star.

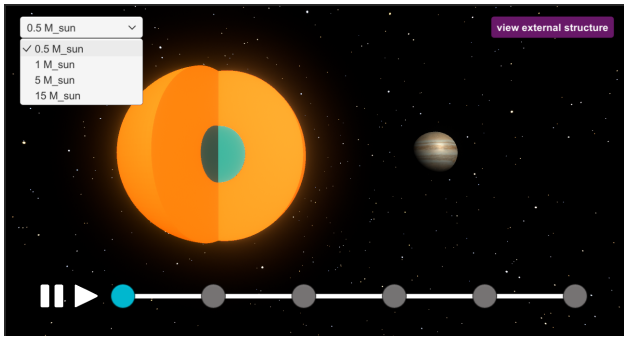

Figure 3: The interface when navigating to a different mass star. In this case, star with 0.5 M.

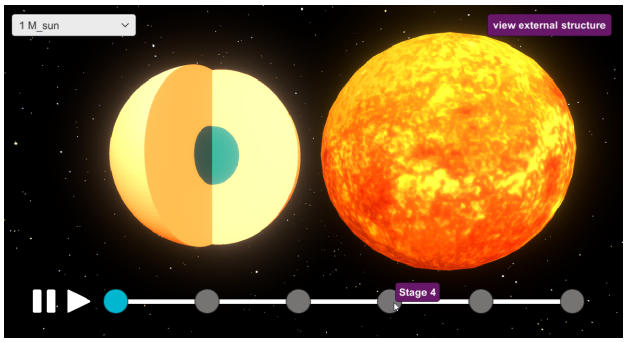

Figure 4: Hovering over objects in the interface provides the user with more information/instructions.

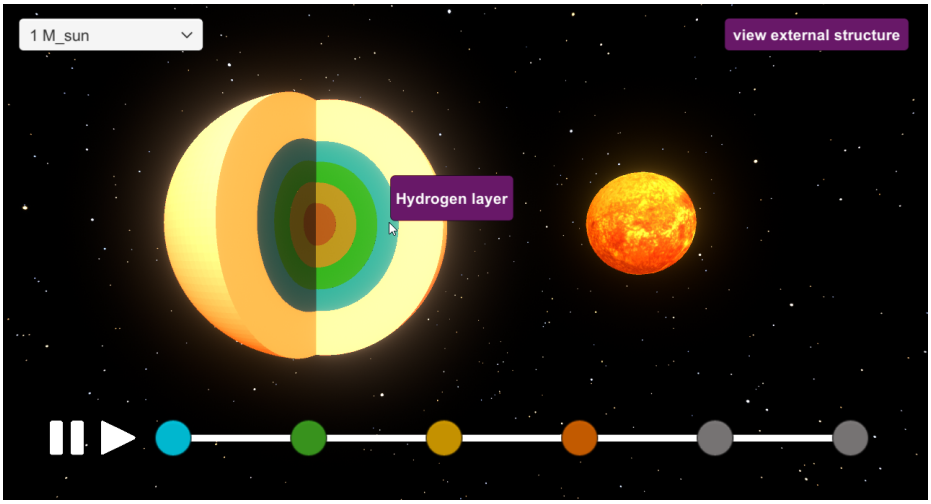

Figure 5: Hovering over layers of the star with a computer mouse gives users more information on that specific layer.

1.2 Instructions

This user evaluation focused on the functionality of the interface and its interactions. The user evaluation also focused on the way the data of the stars was presented in the program. There were two approaches for the user evaluation because the user evaluation was split into two evaluation groups.

1.2.1 Evaluation Group 1

This group mainly consisted of people without any astronomy knowledge. We sat down with one person at the time to give them a total of seven short tasks. These tasks needed to be executed by the user in order to measure the usability score of the program, and to measure the usability attributes of the usability specification (Figure 12). The users were asked to perform the following tasks:

1. "Start the animation of the life cycle of the star."
2. "Pause the animation of the life cycle of the star."
3. "Stop the animation of the life cycle of the star."
4. "Navigate to the fourth stage of the star."
5. "What is the green layer supposed to represent? Try to find out."
6. "View the external structure of the star."
7. "Navigate to a star with a different mass."

While the evaluators performed these tasks, the seconds it took for a user to execute a certain task were timed. In addition, the number of mistakes a user made during the execution were counted.

After performing these tasks, the evaluators had to fill in a survey to evaluate the usability of the program. This was done with the System Usability Scale (SUS) [1]. A template was used for this (Figure 11) [2]. The results of this survey can be found in section 2.

1.2.2 Evaluation Group 2

This group consisted of astronomy students. Since we were only able to reach these users online, they had different instructions than the users in user group 1. To do this, the program was build and published on itch.io [4].
In this approach, the evaluation focused more on the way the data was visualised. A short instruction was given digitally to every user in this group. The users were asked to "play" with the program. They were also asked to fill in a short survey to give feedback on the program, similar to the first user evaluation. Additionally, they were also given the possibility to give suggestions on other features that would make the interface easier to use and understand. The results of this survey are shown in section 3.

2 Results Evaluation Group 1

To measure the system usability of the program, the users had to fill in a survey. The users had to rank each statement on a scale of 1 to 5, where 1 means "strongly disagree", and 5 means

"strongly agree". Figure 6 shows the results of this survey. The rows show the statements that were given in the survey. The lowest score that was given is displayed in the "Lowest Score" column, while the highest score is displayed in the "Highest Score" column. The last column shows the average score of all users in this user group. The calculated System Usability Scale (SUS) scores [3] are displayed in the last row "SUS score". The overall calculated SUS score is 88.5. This means that our program has high usability according to the users in evaluation group 1, since A SUS score that is above 68 is considered to be above average [3]. Even the lowest calculated SUS score is above average.

SUS survey	Lowest Score	Highest Score	Average Score
I think that I would like to use this system frequently.	3	5	3.8
I found the system unnecessarily complex	1	2	1.5
I thought the system was easy to use.	4	5	4.5
I think that I would need the support of a technical person to be able to use this system.	1	2	1.2
I found the various functions in this system were well integrated.	4	5	4.2
I thought there was too much inconsistency in this system.	1	2	1.2
I would imagine that most people would learn to use this system very quickly.	4	5	4.7
I found the system very cumbersome to use.	1	2	1.3
I felt very confident using the system.	3	5	4.4
I needed to learn a lot of things before I could get going with this system.	1	2	1.2
SUS Score	72.5	97.5	88.2

Figure 6: Scores of the system Usability Scale

In addition, the users were also asked two other questions to provide us with feedback.
We asked the users "Did you understand the functions of every part of the interface?". The users gave the following responses:

- "Yes."
- "Yes, I did."
- "At first I didn't realise that hovering over the layers gave more information."
- "The pause button switched to a play button and then there were suddenly two play buttons? That was a bit confusing."
- "The shrinking function was random and confusing to me. I did not know that the star on the right side was supposed to be a reference star."
- "I didn't understand how to read the information of the layers."
- "I did not understand the difference between the different masses, but the rest was very clear."

- "The scale reference wasn't clear to me."
- "Trying to get information on the layers was kind of hard for me to figure out at first."

We also asked the users "Do you have suggestions to improve the clarity of the interface?" The users provided us with the following suggestions:

- "The play and stop buttons were kind of confusing. I'd suggest using a single button for play and pause, and then a different button for stop."
- "Don't switch the pause button to a play button but just keep the pause icon and then maybe press the play button next to it to start playing it again. Or change the pause button to a different icon."
- "Make sure the user knows the star on the right side is a reference for the size of the star on the left. Also add something to clarify how you change the size of the star because it looks like a random tab for me."
- "Add text to the dropdown menu so it is more clear what that menu is for, a single play-pause button and a stop button, and I think it would be more clear if there was some sort of hovering spot on the layers that invites you to hover of the layers. This 'spot' could be like a small dot that lights up when you hover over it."
- "The dropdown button on the left could be made more clear."
- "Add time that has passed to each stage."
- "For the information of the layers: you could create something that makes it more clear that hovering over the layers, provide you with more information."

To get more information on the usability of the interface, we also measured the number of errors, and time it took for the users to complete the given tasks. These tasks are specified in section 1.2.1. The results are shown in Figure 7. This table shows the number of errors each user made during each tasks, and the operation speed (in seconds) of the users during each task. The two bottom rows show the average amount of time it took to complete each task.
In the design document that was created for the production of the program, a usability specification was created 12. This specification contains usability attributes with the lowest and highest number of allowed errors, and with the lowest and highest amount of operation time in seconds. Additionally, the target values we set for these attributes are also shown. Figure 8 shows the first half of the usability specification with the measured results. The set scores are marked with blue, while the measured values are marked with green. For the measured number of errors, the maximum number of errors made by the same user are displayed. For the operation time, the average operation time for the specific task is displayed. As shown in Figure 8, the average amount of time it took for the users to navigate to a specific stage of the life cycle was 1.3 seconds. The average amount of time it took for the users to change the mass of the star was 2.8 seconds. Both operation speeds were below the minimum allowed value (6 and 8 seconds, respectively), and were better than the target value we set for both attributes. The same applies for the error tolerances of both operations. For navigating to a specific time, only 1 error was made, and for changing the mass of the star, no errors were made. Again, for both usability attributes, the measured values were lower than the minimum value. Additionally, the number of errors was better than the error tolerance goal that was set. Therefore, it can be said that the measured usability attributes satisfied our predictions and goals. Since these two interactions

are the main interaction features of this program, we can say that the program seems to be usable.

		Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7
User 1	Time (s)	0.5	0.5	0.5	1	1	1.5	3
	Errors	0	1	0	0	0	0	0
User 2	Time (s)	1	1	0.5	3	2	3	1
	Errors	0	0	0	1	0	1	0
User 3	Time (s)	1	0.5	0.5	1	1	4	2
	Errors	0	0	0	0	0	0	0
User 4	Time (s)	0.5	0.5	0.5	0.5	0.5	2	2
	Errors	0	0	0	0	0	0	0
User 5	Time (s)	1	0.5	0.5	1.5	2	2	3
	Errors	0	0	0	0	0	0	0
User 6	Time (s)	0.5	0.5	0.5	2	1.5	2.5	2
	Errors	0	0	0	0	0	1	0
User 7	Time (s)	1	1	0.5	1	3	1	1.5
	Errors	0	1	0	0	1	0	0
User 8	Time (s)	0.5	0.5	0.5	0.5	2	1	1.5
	Errors	0	0	0	0	0	0	0
User 9	Time (s)	2	1	3	2	6.5	5	8
	Errors	0	0	1	0	2	0	0
User 10	Time (s)	0.5	0.5	1	1	3	2	3
	Errors	0	0	0	0	0	0	0
User 11	Time (s)	0.5	0.5	0.5	1	2.5	1.5	4
	Errors	0	0	0	0	0	0	0
Average number of time		0.8	0.6	0.8	1.3	2.3	2.3	2.8
Total number of errors		0	2	1	1	3	2	0

Figure 7: The number of execution errors a user made per task, and the operation speed of the users per task.

Usability attribute	Measurement	Value to measure	Min	Goal	Max	Measured
Operation speed	Navigate to specific time	Time (s)	6	3	1	1.3
Operation speed	Change mass of star	Time (s)	8	4	1	2.8
Error tolerance	Navigate to specific time	Number of errors	2	1	0	1
Error tolerance	Change mass of star	Number of errors	2	1	0	0

Figure 8: First part of the usability specification with the lowest acceptable scores (Min), the best possible scores (Max), the target scores (Goal), and the measured values for these attributes.

3 Results Evaluation Group 2

The second evaluation group was asked to fill in a short survey to give feedback on the program. They were also asked to rate the suitability, educational value, clarity, and satisfaction of the program, and its interface. The results are summarised in Figure 9. As shown in this figure, all measured scores are higher or equal to the target ('goal') score. This means that the interface satisfies the set expectations of the usability attributes.

Usability attribute	Measurement	Value to measure	Min	Goal	Max	Measured
Suitability	Questionnaire	Average score	5.5	8	10	8.2
Satisfaction	Questionnaire	Average score	5.5	8	10	8
Educational value	Questionnaire	Average score	5.5	8	10	8.2

Figure 9: Second part of the usability specification with the lowest acceptable scores (Min), the best possible scores (Max), the target scores (Goal), and the measured values for these attributes.

3.1 Suitability

We asked users: "Is the way the visualisation is presented suitable for this kind of information, or would a different visualisation work better?".

The users gave suitability an average score of 8.2/10, and provided the following suggestions:

- "Make the information of the layers more clear, and add some sort of time indication."
- "A timeline would maybe also be suitable. (Maybe this can be added at the stages below with some small pictograms or text)."
- "Maybe add the labels not just when hovering over a certain part of the animation, but always. Especially when the sizes of orbits are flying by, one cannot really see it. Maybe also clarify that the picture on the right is the Scaling picture, I first thought that the sun was shrinking because I did not fully understand what was happening. Maybe also add a legend for the colours."

3.2 Educational value

We asked users: "Do you think this visualisation is valuable as an educational tool for astronomy students?".

The users gave educational value an average score of 8.2/10, and provided the following suggestions:

- "Add more information to the layers."
- "I'd suggest adding more information to the stages."
- "I think some extra information would be great. Now the star is growing and shrinking, but I have no idea why and when. At some point Jupiter and Mercury come into the screen, but it is unclear to me what they are doing there (even when i hover over it with my mouse). Due to the moving of the comparing objects your focus shifts to them, while the main object of focus is the star. I would have find it clearer if the star were more in the center of attention and is explained to be the evolving star."
- "Maybe add a timescale, or more details to the "Stage X". Maybe also not completely replace Jupiter with the sun but show them side to side. Otherwise one also has to know how the sun scales w.r.t. Jupiter and the orbits of the inner planets."

3.3 Clarity

We asked users: "Did you understand the functions of every part of the interface?".

The users provided us the results shown in Figure 10, and provided the following suggestions:

- "The external structure felt as if it did not add too much. At first I was also slightly confused as to how to think of the size of the star; it was not absolutely clear to me that when the right hand side star grows, the 'real' star actually is shrinking and vice versa."
- "The hovering over function of the layers wasn't clear to me at first."
- "It took me a while that I could hover my mouse over to objects to read what they are doing there."
- "Maybe add a little box on every layer on which you have to click on in order to see the information of the layer. Now it is quite unclear what you should do to find the information of the layers."

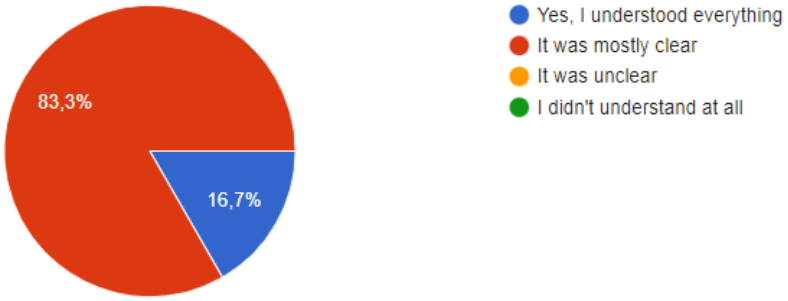

Figure 10: A pie chart of the results about the clarity of the interface.

3.4 Satisfaction

We asked users: "How would you rate using this interface overall?".

The users gave satisfaction an average score of 8/10, and provided the following suggestions:

- "The interior of the star is a bit unclear. The yellow has no identification while the blue is Hydrogen (I do not know if a layer is missing or the labels are wrong.) A suggestion could be to give a small pop up screen at the start that explains some of the functionalities and what we are about to see, otherwise you go in without any context and only see some moving objects)."
- "Make it possible to click ahead in the phases and start the simulation from there."

3.5 The research question

The research question of this project is "To what measure will an interactive animation help users understand what reactions take place in a specific layer of an evolving star, and help them understand what the life cycle of the star looks like?". To get a better indication on how to answer this question, we asked the evaluators: "To what extend do you think this program will

help you study and memorize the life cycle of a star?"

The users gave the following suggestions:

- "I think that when every feature works and more information is added, it will definitely help me memorize it better."
- "At the point of filling in the survey, the program was not complete and not all stages were added. This is a clear visualisation but as of yet, I did not learn anything. I had to use my own knowledge to know what is going on. More information in text that cannot be represented as an animation would have helped me."
- "It will help me get a better (basic) picture of what the life cycle looks like. This helps me with studying because I always have a hard time imagining it."
- "This is more fun than learning things from a book. I also prefer learning things via visual representations because this way I actually see the process and that helps me with understanding it."
- "I think not much helping memorizing it, but it can help a lot for getting a feeling of the time and physical scale of these things."

4 Analysis

When looking at the results of the evaluation done by user group 1, it can be said that the program satisfies the desired usability. Overall, the main interactions of the system were clear. The measured operation time, as well as the number of errors users made, were low. In addition, the SUS score of the program is high, which means that the scores that were given during the evaluation were positive. Moreover, the feedback given by this user group also shows that almost every function of the interface was clear. However, there are some interactions that caused some confusion to certain evaluators. The hovering function that provides the user with more information on the layers of the stars, were often mentioned as something that was unclear. Another thing that was mentioned, were the reference objects for the scale of the star, and the differences between the stars with different masses. It was also mentioned to explain that the drop down menu in the interface is to navigate to a star with a different mass. Additionally, someone also said that the play, pause, and stop buttons were confusing.

The survey of the evaluation done by user group 2, resulted in high suitability, educational, and satisfaction values. The evaluators also concluded that most parts of the interface were clear. However, they also provided us with feedback on how to make certain interactions better, since there still was some confusion about certain aspects of the interface. They also gave suggestions on how to improve the educational part of this program. Again, the hover function to gain more information on the reference objects, and the layers, did not work optimally according to some of the evaluators. Also, the users would like to see some sort of time indication to know how much time has passed during each stage. In addition, more information about the layers, temperatures, and size of the star was often requested. This is something that we were already planning to do, but because of the feedback we now have a more clear idea on how to implement that. One of the evaluators also mentioned adding a little box to each layer. This way the users will know that there is an interaction possible regarding the layers to get more information. Another evaluator suggested creating a pop up screen at the start that explains the functionalities and visuals of the program.

For most part, the evaluators in the second evaluation group agreed that this program could help them with getting a better overview of the life cycle of stars with different masses because of the visual representation. In order for this to apply, more information needs to be added to the program, and some features of the interface need to be improved.

5 Action Points

These are the main points we have taken away from this user evaluation and will implement for the final product:

- Add more information when hovering the cursor over the layers of the star, and make it more clear that the users are able to interact with the layers. This can be done by adding a small box where the user can hover over with its mouse.
- Add an information panel to the interface which explains how the program works and what assumptions were made regarding physics.
- Simplify the start, stop, and pause buttons by creating a single pause and play button, and by creating a separate stop button to stop the animation.
- Make it more clear that the objects on the right are reference object, and make sure the users understand the scaling reference.
- Add text to the drop down menu, for changing the mass of the star, to make it clear what the menu is meant to do.
- Add a time indication to the stages of the life cycles.

References

[1] Brooke, J.: SUS-a quick and dirty usability scale. Usability Eval. Ind. 189(194), 4–7, 1996

[2] Meiert, J. O.: Revitalizing SUS, the System Usability Scale., 4-5, April 23. 2007, retrieved on December 13. 2021 from https://meiert.com/en/blog/revitalizing-sus-the-system-usability-scale/

[3] Thomas, N.: How To Use The System Usability Scale (SUS) To Evaluate The Usability Of Your Website., Retrieved on December 16. 2021 from https://usabilitygeek.com

[4] https://itch.io/

Appendix

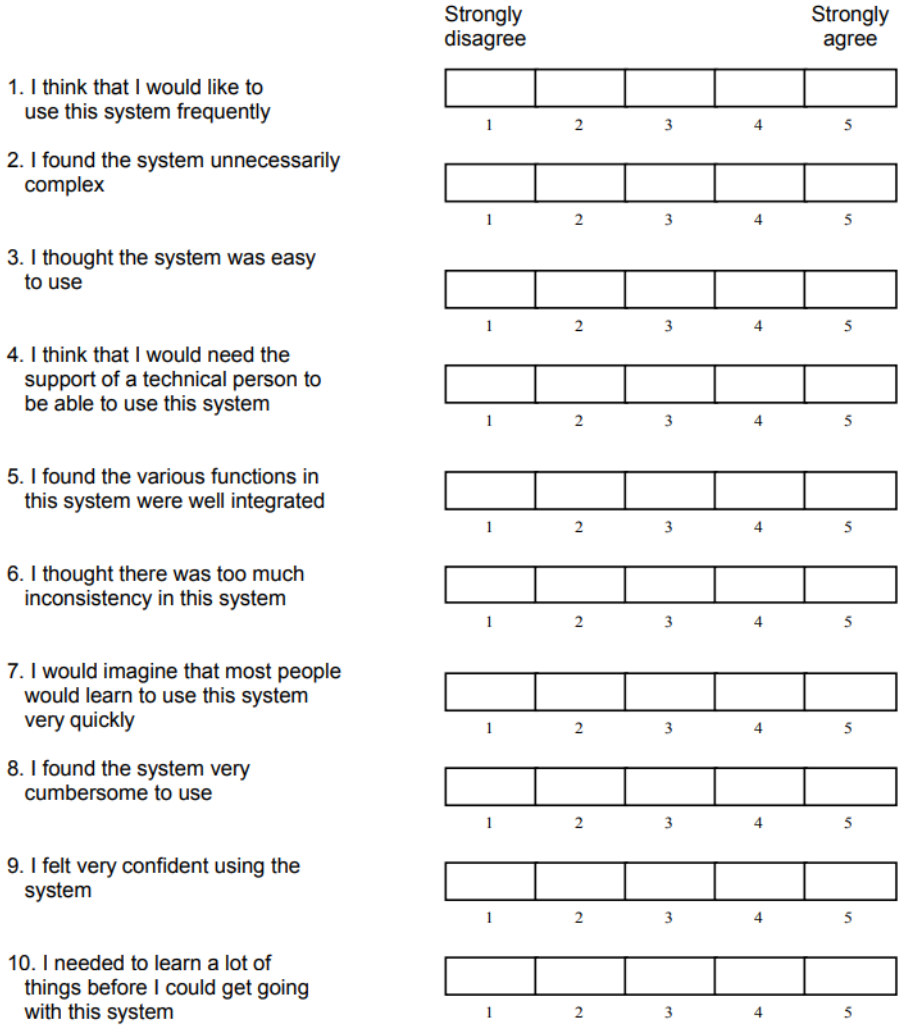

Figure 11: System Usability Scale

Usability attribute	Measurement	Value to measure	Min	Goal	Max
Operation speed	Navigate to specific time	Time (s)	6	3	1
Operation speed	Change mass of star	Time (s)	8	4	1
Error tolerance	Navigate to specific time	# errors	2	1	0
Error tolerance	Change mass of star	# errors	2	1	0
Suitability	Questionnaire	Avg score	5.5	8	10
Learnability	Questionnaire	Avg score	5.5	8	10
Satisfaction	Questionnaire	Avg score	5.5	8	10
Educational value	Questionnaire	Avg score	5.5	8	10

Figure 12: Usability specifications. 'Min' is the lowest acceptable score, 'Goal' is our target score and 'Max' is the best possible score.