

Staff Associate Evaluation Task: Book Content & Beliefs Survey

Ethan R. Feldman

Author Note

Correspondence concerning this report should be addressed to Ethan R. Feldman. Phone: 203-362-9387. Email: etfeldman0@gmail.com. Thank you to Dr. Fong and Dr. Gui from the Columbia Business School for taking the time to review this report. Pages 2-5 (4 pages) encompass the report, while pages 7-15 (appendix) contain supporting visuals.

Staff Associate Evaluation Task: Book Content & Beliefs Survey

The data is from a study titled “Quantifying Emotion” run by Professor Hortense Fong at Columbia University. The goal of the study was to understand readers’ story expectations. Participants read two chapters from each of two fictional stories, then rated their emotional reactions—measured using valence and arousal—for each chapter. Valence measures how positive or negative the participant feels and arousal measures how high energy or low energy the participant feels. The data includes additional behavioral data, attention checks, genre preference questions, and free-text responses. A summary of columns is found in appendix A1.

A Python jupyter notebook and JMP workbook were used for this assignment. My STATA license is currently expired, so the analogous JMP program was used. Throughout the analysis, I consulted online forums like StackExchange and genAI tools like ChapGPT to help brainstorm, debug, and interpret results. LaTeX was used to formalize and write the report. This assignment was completed over several days in approximately 8 hours—a breakdown in appendix A1.

Data Preprocessing

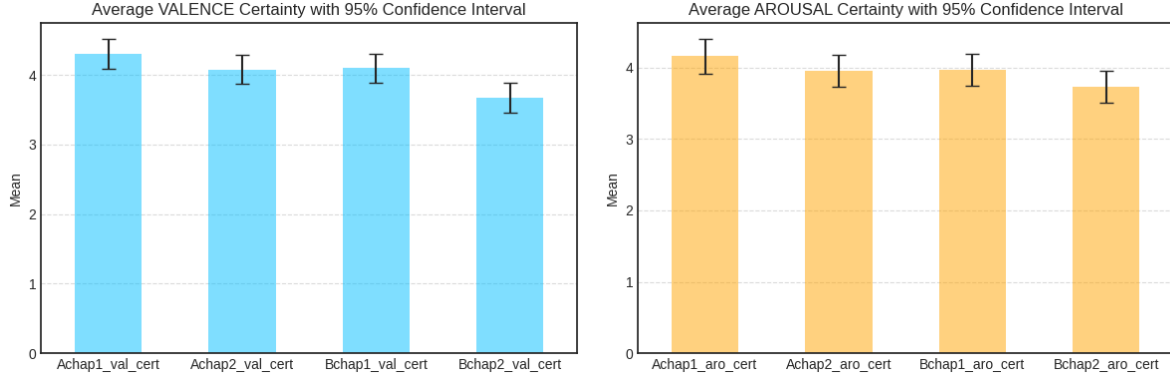
Assuming most 0’s in that dataframe aren’t meant to be null values, the data was mostly complete and needed minimal cleaning. Row 6 and 200 were combined into a single response. Both rows were missing the same data from each other, this was most likely the same response but a data input error. Row 7 contained an "I" in a numeric column, which I inferred to be a 1. Row 1 contained all survey prompts, which I transferred to a different file to keep as a reference. Column 'totalOffTask' was converted from negative to positive numbers. A handful of computed columns were created to support analysis and described in appendix A1.

Analysis

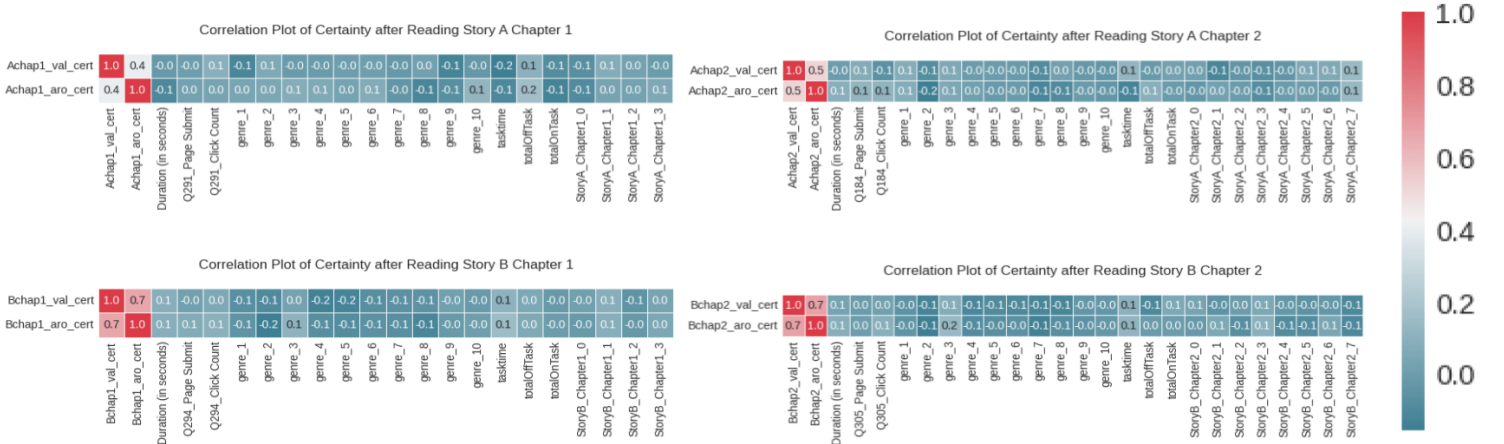
To start, I uploaded the assignment objective and variable summary in ChatGPT to get recommendations. The output suggested looking at belief responses, certainty ratings, and surprise ratings [A2]. My intuition chose to focus on certainty ratings. By focusing on this variable we can begin to understand how users determine confidence in their story expectations. Figure 1 highlights participant certainty scores declining as the survey progresses in both valence and arousal [anova reports A3 and A4].

Figure 1

Plots of diminishing valence and arousal certainty over chapters. Certainty rating decreases from chapter 1 to chapter 2 and overall decreases throughout the survey.

**Figure 2**

Correlation heatmaps for valence and arousal certainty for each chapter.

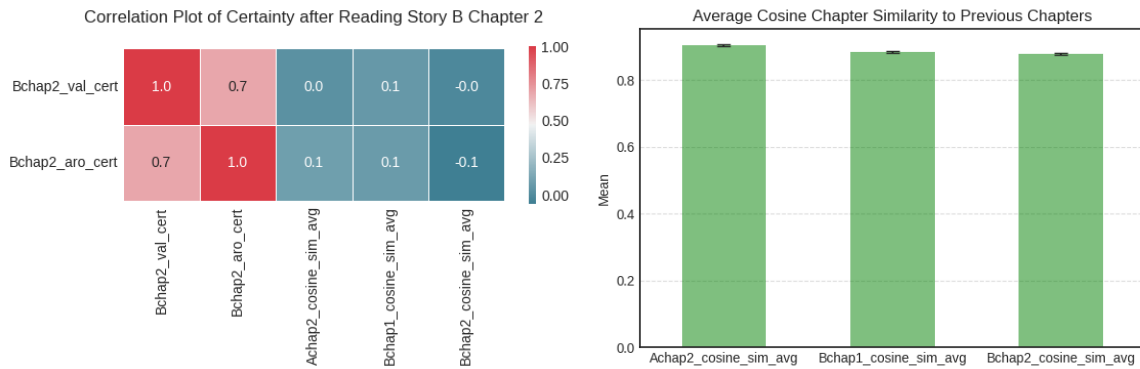


My initial thought was that the chapters were the culprit of the decline. Maybe certain texts were less engaging for the participants. Perhaps genre preferences or other general summaries could also influence the decline. For valence and arousal certainty, I recorded the correlations between genre, chapters (1-hot encoded), and other user metrics shown in figure 2.

Initially, we see valence and arousal certainty ratings are highly correlated with each other. This is intuitive, but not an interesting finding. There is slight evidence for a couple of correlations: A_Chapter1 arousal and total time distracted (offTaskTime), A_Chapter1 valence and total time spent on the key task (taskTime), A_Chapter2 or B_Chapter1 arousal and

Figure 3

Diminishing cosine similarity scores and correlation heatmap of similarity scores by story B chapter 2.



number of fantasy books read (genre_2), B_Chapter1 valence and number of mystery or thriller books read (genre_4 and genre_5), B_Chapter2 arousal and number of biographies read (genre_3). Nothing is consistent or strongly correlated. However, given more time, this leaves me with more questions to explore further.

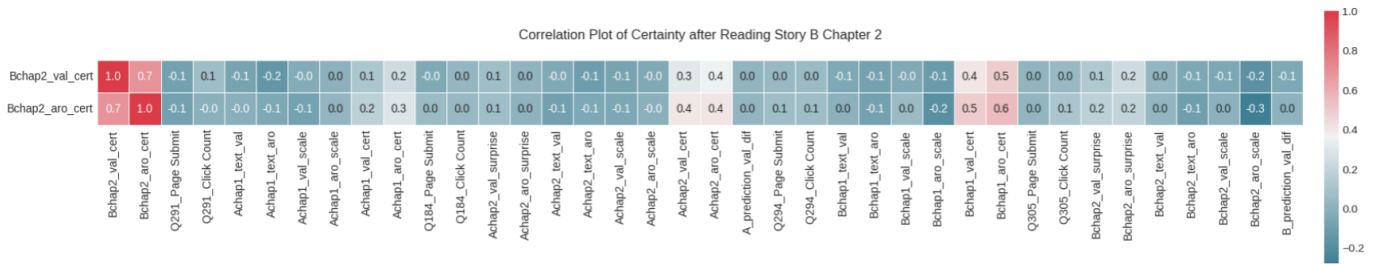
To potentially rule out the effect of the text, I wanted to see if the chapters were similar to the previous ones. Participants read one chapter after the other, story B after story A. The hypothesis is that if chapters become progressively dissimilar to each other, then this could cause confusion leading to drop in certainty. Consulting ChatGPT, I utilized a cosine similarity function that measures the average similarity between the current and previous chapters. Chapters were transformed into embeddings using a BERT tokenizer, and those embeddings were used to calculate cosine similarity. Figure 3 shows that progressively throughout the survey the average similarity between the chapters decreases. However, consulting a correlation chart, there is not enough evidence to support a correlation between the cosine similarities and certainty in valence or arousal. Given more time, a more sophisticated model might better understand this phenomenon.

I liked the idea that previous information could influence certainty. We have current and previous self-ratings and surprise ratings. For each chapter, I plotted the correlations between the valence and arousal certain to the current and previous valence/arousal ratings. All plots found in appendix A5. Looking at the final correlation plot [4] we see evidence for a compounding certainty score effect. The plot suggest that participants are aligning their later certainty

responses more closely with their previous ones. Stepwise regression reports support this hypothesis by showing diminishing p-values as we get farther from the current certainty rating [A6 and A7]. However, relatively low adjusted R-squared values (.18 for valence certainty) and (0.37 for arousal certainty) suggest more going on. Surprisingly, whether their arousal or valence ratings match their predictions (*_prediction_(val/aro)_dif) doesn't show evidence of effecting prediction certainty. Surprise in rating difference and previous chapter valence or arousal rating shows slight correlation, but further analysis is needed to explain this.

Figure 4

Correlation of valence and arousal (for next chapter) after reading story B chapter 2.



Conclusion

Evidence suggests participants align their valence and arousal certainty with prior predictions, with a stronger influence from recent predictions than distant ones—implying user confidence depends on recent experiences/information. Given 8 more hours, I'd explore: (1) applying a machine learning model with explainability, and (2) studying the role of user attention.

Preliminary analysis indicates other influences on certainty, like surprise and genre. Within 2 hours, I could train a Scikit-Learn model with the additional identified features, and in another 2 hours, apply explainability frameworks to identify influential factors and/or effects—though this may exceed 6 hours depending on complexity. In the final 2 hours, I'd design a follow-up survey to test whether certainty diminishes over time during a longer story. Possibly we could explore disengagement/fatigue which is common in longer studies, or find an opposite effect where users certainty might increase after a percentage threshold of text is read.

Previous studies highlight the importance of a strong “hook” (Folk and Gibson, 2001; Knight et al., 2024; Pieters and Wedel, 2004) to reel people in, but to retain information, I suggest presenting your most important content at the end.

References

- Folk, C., & Gibson, B. (2001). *Attraction, distraction and action: Multiple perspectives on attentional capture* (Vol. 133). Elsevier.
- Knight, S., Rocklage, M. D., & Bart, Y. (2024). Narrative reversals and story success. *Science Advances*, 10(34), eadl2013.
- Pieters, R., & Wedel, M. (2004). Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. *Journal of marketing*, 68(2), 36–50.

Appendix
Supporting Figures and Tables

Figure A1
Breakdown for time spent on this report. Allowances are not exact and frequently overlap.

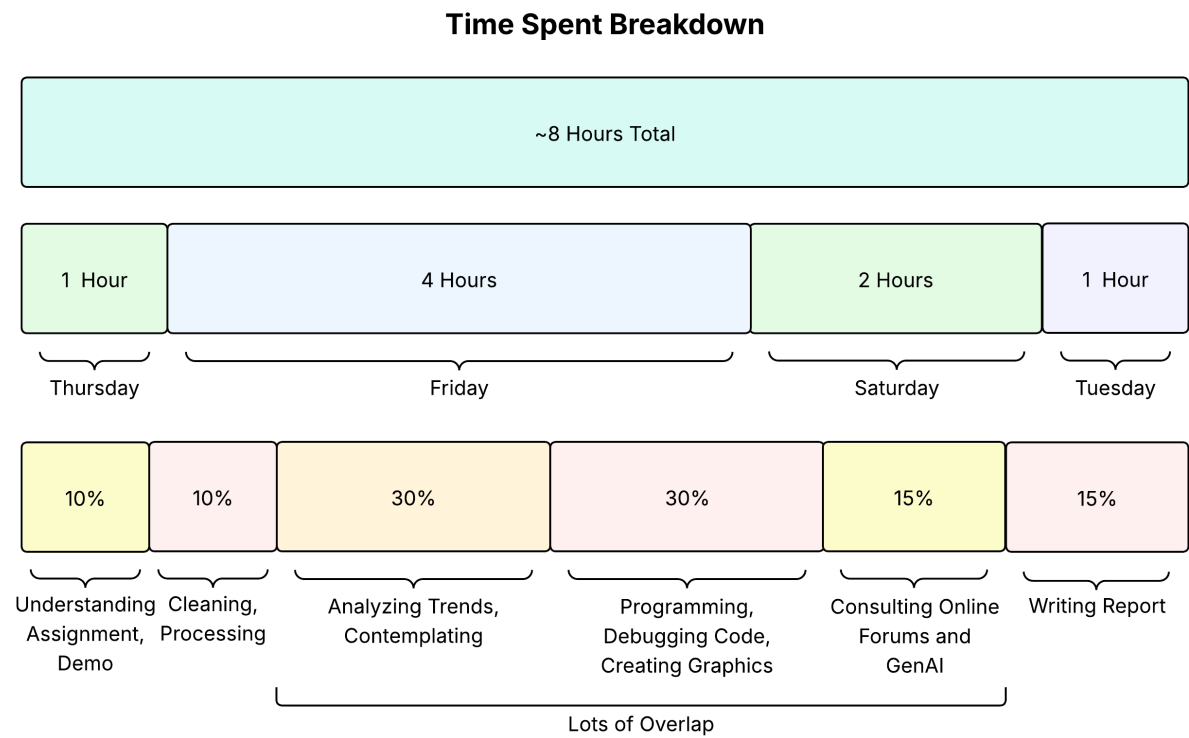


Figure A2

ChatGPT recommendations for prominent variables.

Out of these variables, which might be insightful for understanding user story expectations?

Happy to help! Could you share the list of variables you're referring to? Once I see them, I can suggest which ones are most insightful for understanding user story expectations.

Survey Variable Documentation

Metadata and Timing

StartDate: Timestamp when the respondent started the survey.

EndDate: Timestamp when the respondent ended the survey.

Duration (in seconds): Total duration of the survey in seconds.


RecordedDate: Timestamp when the response was recorded.

RespondId: Unique respondent ID.

Page Interaction Timing


Q291_First Click: ↓ click time for the initial page.

Q291_Last Click: Last click time for the initial page.

 **Belief-Based Variables (Direct Expectations)**


These are *explicit indicators* of user expectations:

- `Achap1_belief`, `Achap2_belief`, `Bchap1_belief`, `Bchap2_belief`:
Open-ended beliefs about what will happen next—**direct reflections of narrative expectations**.

 **Affective Appraisal Variables (Emotional Framing of Expectations)**

These reveal **emotional tone** of expectations and confidence:

- `*_val_scale`, `*_aro_scale`:
Valence and arousal ratings—helpful to know how users feel the story is progressing.
- `*_val_cert`, `*_aro_cert`:
Confidence in their ratings shows **how certain** they are in interpreting the story tone, which ties to expectations.
- `*_val_cert_exp1`, `*_aro_cert_exp1`:
Their explanations give deeper **rationale** for expectations.

 **Surprise Variables (Expectation Violation)**

Useful to assess **how expectations were broken**:

- `Achap2_val_surprise`, `Achap2_aro_surprise`, `Bchap2_val_surprise`, `Bchap2_aro_surprise`
— and their corresponding `*_exp1` variables:
These gauge the **gap between expectation and outcome**.

↓

Table A1

Column summary of data collected in the Quantifying Emotion study. Calculated columns are shown at the bottom with accompanying description.

Data Category	Description
Metadata and Timing	Timestamps for when the user started and ended the survey, total duration, and response ID.
Page Interaction Timing	Timestamps for when the user clicked through the pages of the survey.
Story A - Chapter 1 Responses	Valence and arousal ratings after reading chapter 1, predictions for next chapter, and rating reasoning.
Story A - Chapter 2 Responses	Valence and arousal ratings after reading chapter 2, surprise rating, predictions for next chapter, and rating reasoning.
Story B - Chapter 1 Responses	Valence and arousal ratings after reading chapter 1, predictions for next chapter, and rating reasoning.
Story B - Chapter 2 Responses	Valence and arousal ratings after reading chapter 2, surprise rating, predictions for next chapter, and rating reasoning.
Genre Preferences	Number of books the participant has read in various genres in past 10 years.
Attention and Comments	Typed responses from participants summarizing the stories read and opportunity to provide feedback.
Stimuli Text	Text from both stories.
Engagement Metrics	Time spent on and off the survey, total time spent on key tasks, and timestamps when the user switched off task.
Calculated Variable	Description
Averages	Mean rating, certainty, and surprise for valence and arousal.
Prediction Difference	Valence and arousal differences between prediction and actual, average differences.
times_distracted	Count of times distracted (off task).
Chapter Encoding	1-Hot encoding and BERT vector embedding representations of chapters.

Figure A3

ANOVA reports for valence certainty over chapter. The left report contains a means comparison over all chapters. The right report contains the mean comparison blocking for story. Both highlight a decline in user valence certainty.

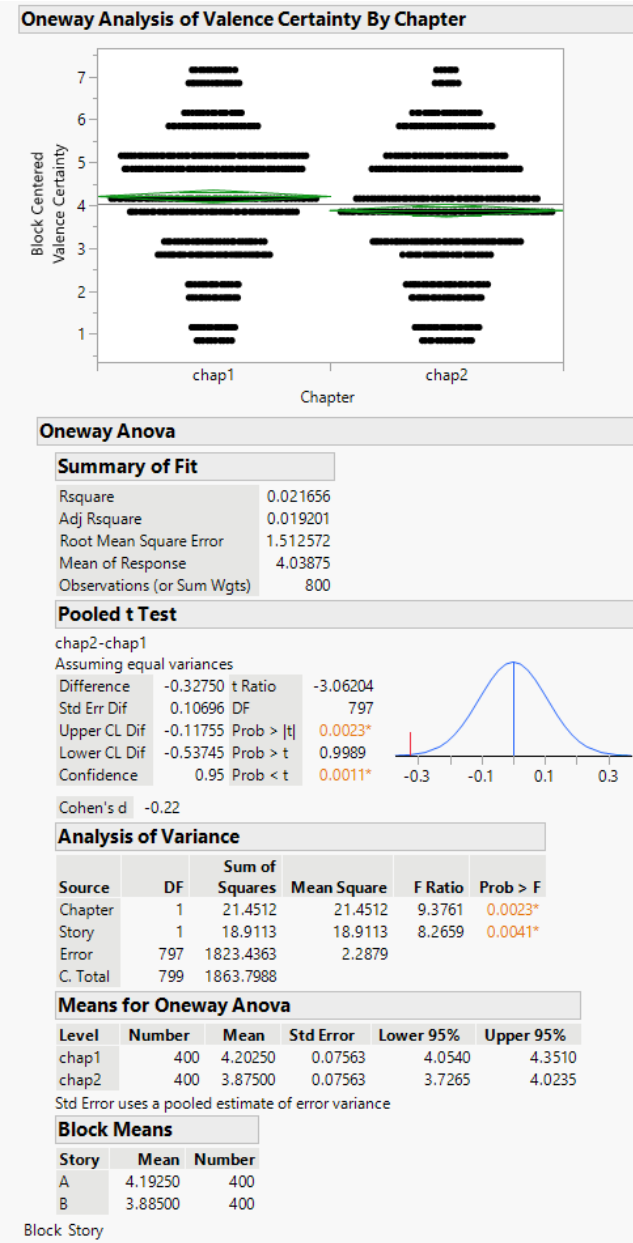
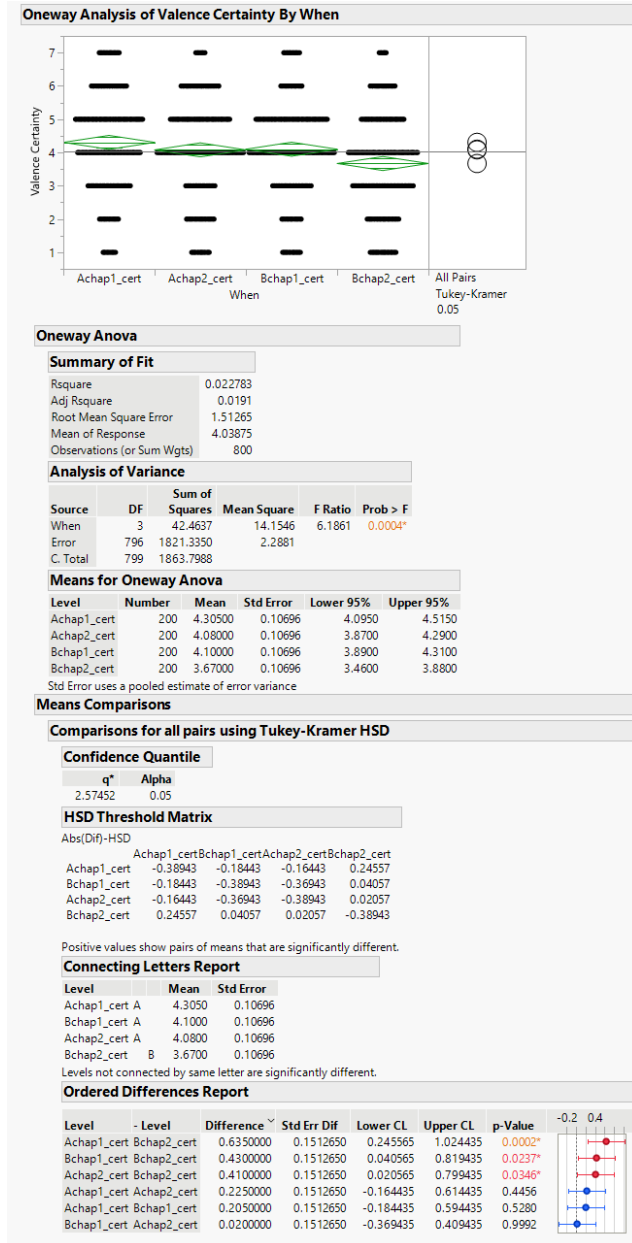


Figure A4

ANOVA reports for arousal certainty over chapter. The left report contains a means comparison over all chapters. The right report contains the mean comparison blocking for story. Both highlight a decline in user arousal certainty.

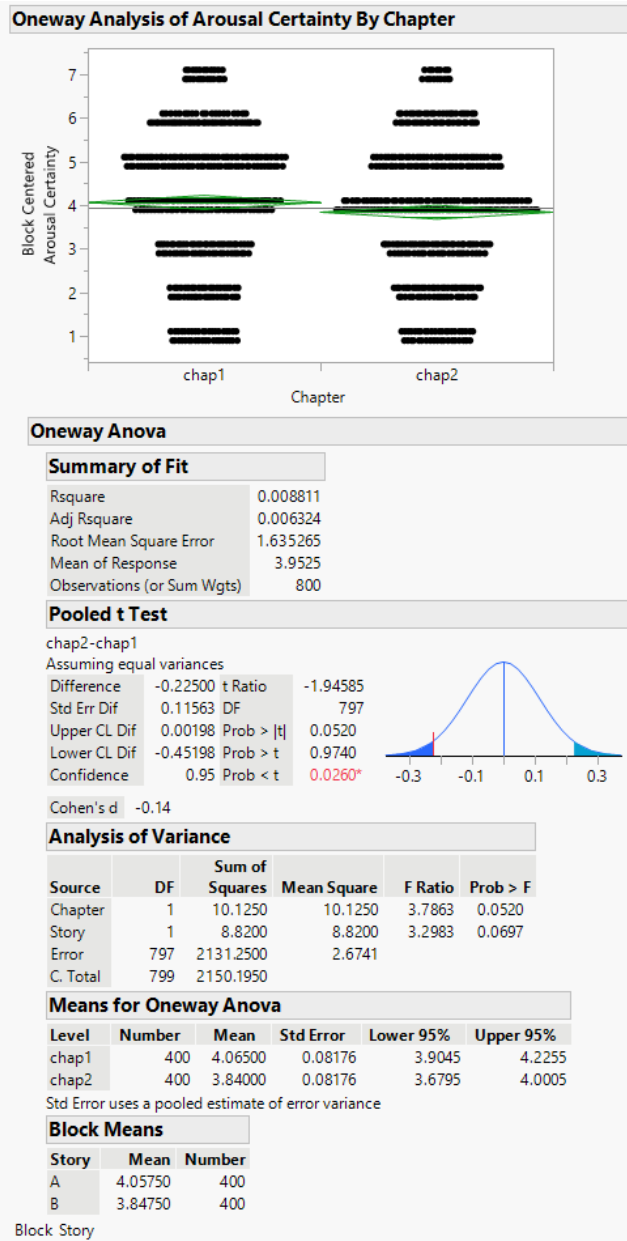
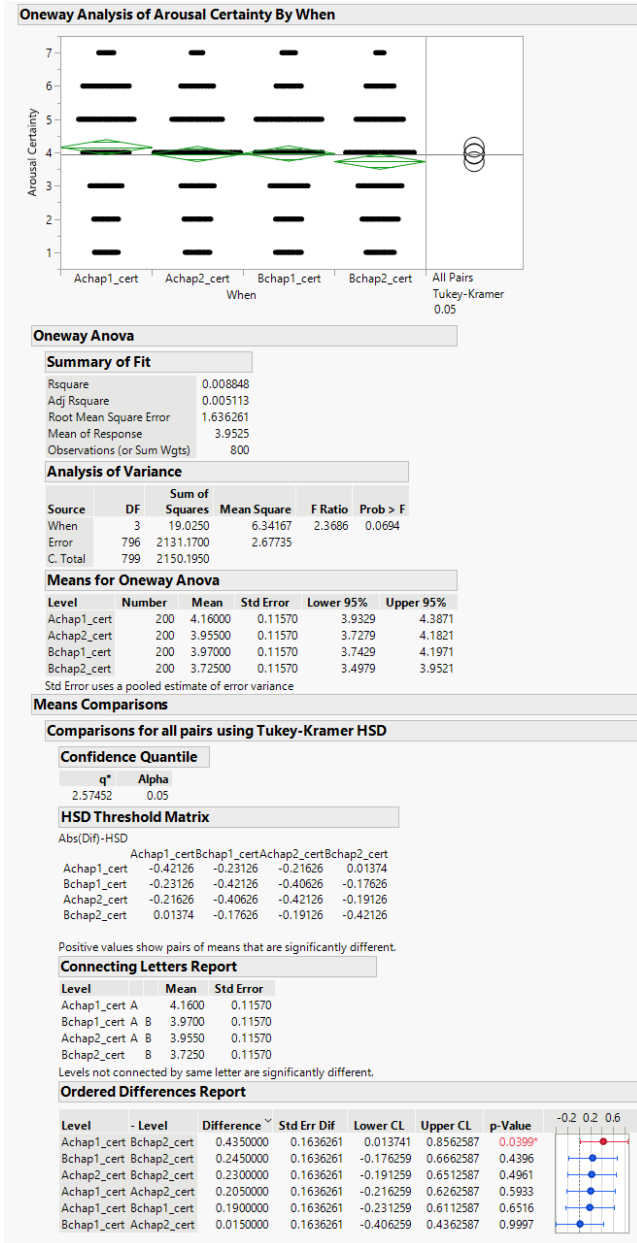


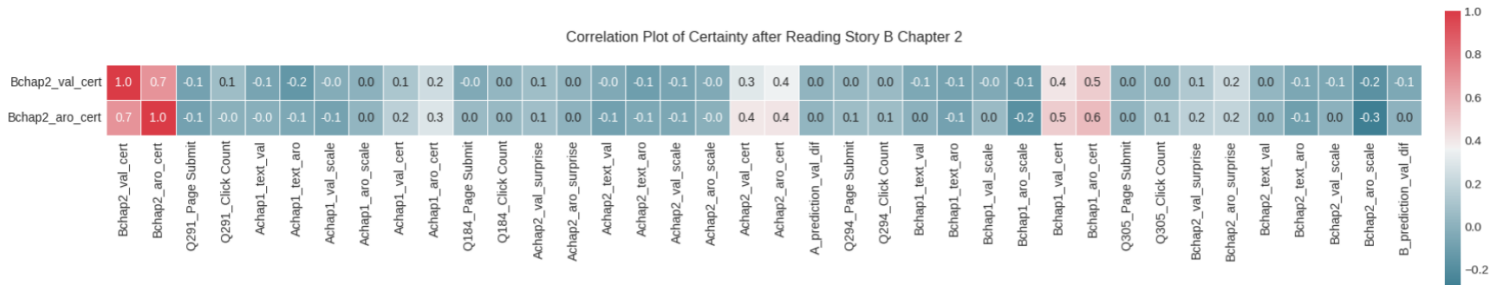
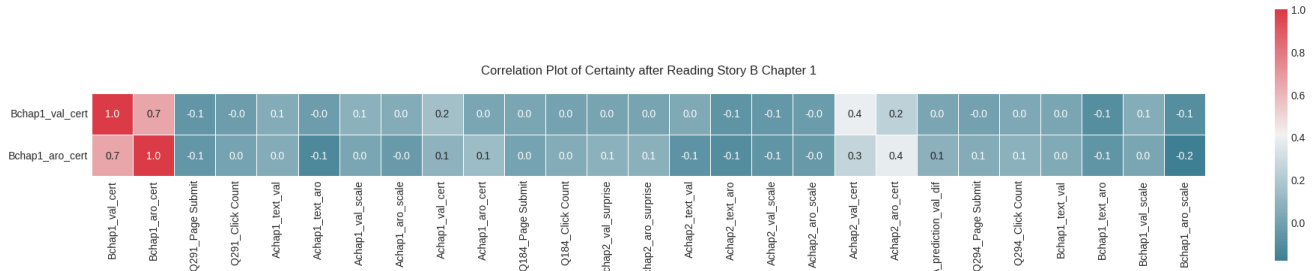
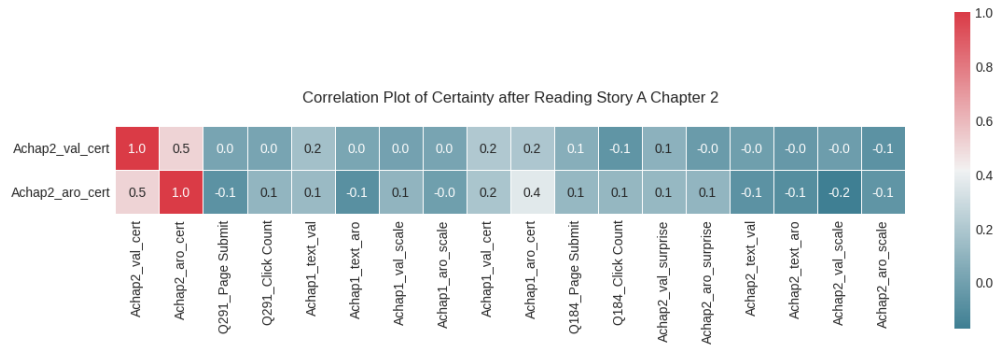
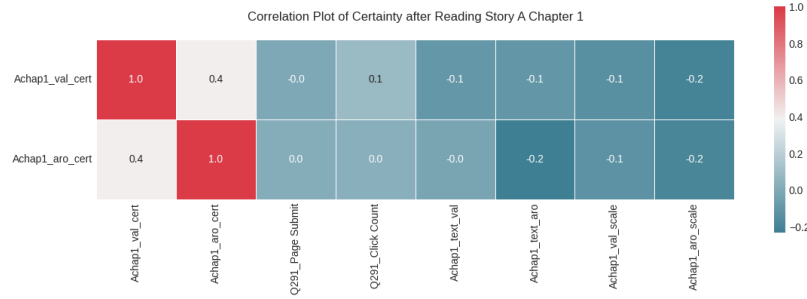
Figure A5*Correlation plots for valence and arousal certainty at each chapter throughout the survey.*

Figure A6

Stepwise regression report for valence certainty over chapter. Simple effect summary table provided.

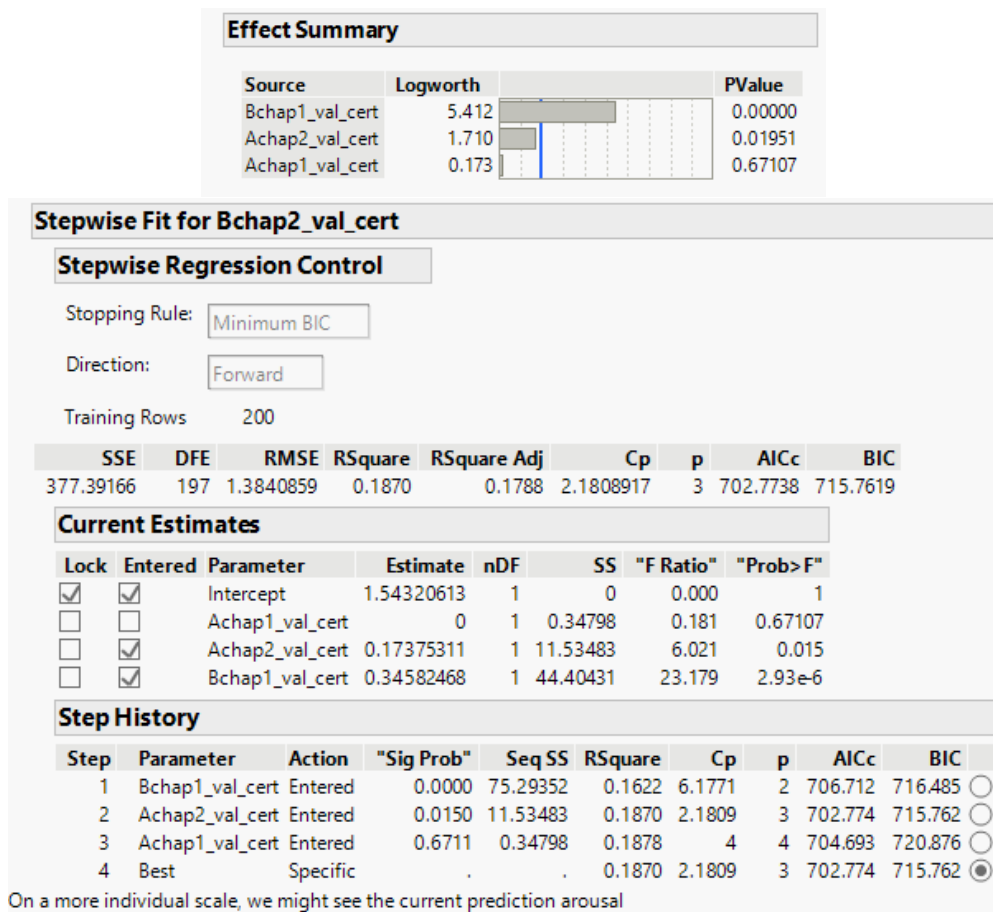


Figure A7

Stepwise regression report for arousal certainty over chapter. Simple effect summary table provided.

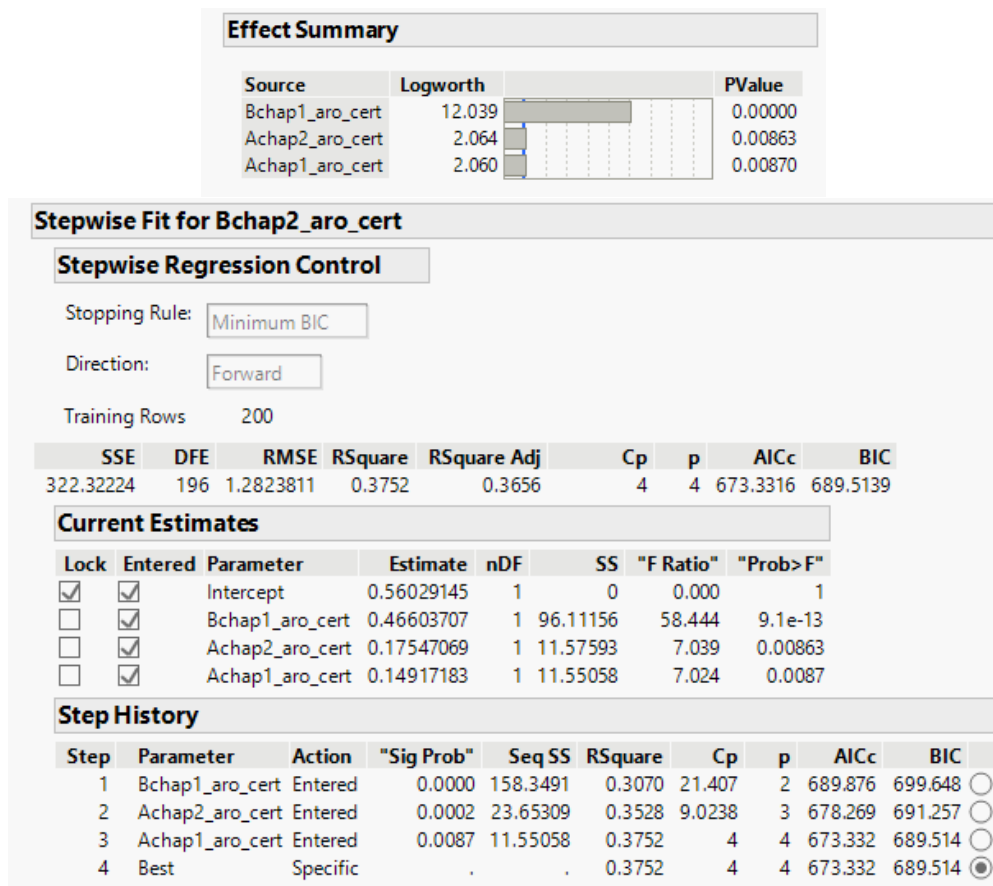


Figure A8

ChatGPT interaction for help with LaTeX formatting.

latex make line spaces smaller for tables

To make line spacing smaller in LaTeX tables, you can reduce the row height using one or more of the following techniques:

1. Use `\renewcommand{\arraystretch}{value}`

This controls the vertical spacing in tables. The default is `1`. Values less than 1 reduce spacing.

```

latex
\renewcommand{\arraystretch}{0.9} % Reduces space between rows
\begin{tabular}{ll}
Header 1 & Header 2 \\
\hline
Row 1 & Data \\
Row 2 & Data \\
\end{tabular}

```

I don't want to change the space in between rows, I want to change the space between the lines in the cells

If you want to reduce the line spacing **within** cells (not between rows), you can use the following methods:

1. Use `\renewcommand{\baselinestretch}{value}` Locally

Wrap your table in a `\begin{group}` and `\end{group}` to change only the text inside the table.

```

latex
\documentclass{article}
\usepackage{array}

\begin{document}

\begin{group}
\renewcommand{\baselinestretch}{0.8} % Reduce line spacing inside cells
\small % Optionally, reduce font size

```