HumSum: A Personalized Lecture Summarization Tool for Humanities Students Using LLMs

Zahra Kolagar
Fraunhofer IIS
Erlangen, Germany
zahra.kolagar@iis.fraunhofer.de

Alessandra Zarcone
Fraunhofer IIS, Erlangen, Germany
Technische Hochschule Augsburg, Germany
alessandra.zarcone@tha.de

Abstract

Generative AI systems aim to create customizable content for their users, with a subsequent surge in demand for adaptable tools that can create personalized experiences.

This paper presents HumSum, a web-based tool tailored for humanities students to effectively summarize their lecture transcripts and to personalize the summaries to their specific needs. We first conducted a survey driven by different potential scenarios to collect user preferences to guide the implementation of this tool.

Utilizing Streamlit, we crafted the user interface, while Langchain's Map Reduce function facilitated the summarization process for extensive lectures using OpenAI's GPT-4 model. HumSum is an intuitive tool serving various summarization needs, infusing personalization into the tool's functionality without requiring personal user data collection.

1 Introduction and Motivation

The educational landscape underwent a significant shift with the advent of e-learning platforms and educational applications, particularly during the global pandemic. The increased reliance on remote learning highlighted the necessity of digital tools to support students' educational needs (Maatuk et al., 2022; Fauzi, 2022; Lynch, 2020; Radha et al., 2020). The surge in online learning applications underscores the critical need for personalized tools capable of meeting diverse student requirements. One-size-fits-all approaches often struggle to accommodate the varied learning paces, preferences, and strengths of individual students (Tetzlaff et al., 2021; Reber et al., 2018). An essential aspect lies in creating educational applications and tools that go beyond generic content delivery, ensuring tailored experiences that resonate with each student's unique learning trajectory. Personalization not only enhances academic outcomes but can also serve as a motivational factor, fostering student enthusiasm and persistence in learning endeavors (Maghsudi et al., 2021a; Sadovaya et al., 2016).

The evolution of Large Language Models (LLMs) represents a recent advancement in artificial intelligence, with the possibility of significantly influencing various facets of education. Models such as pre-trained GPT-3/4 (Koubaa, 2023; OpenAI, 2023), Llama models (Touvron et al., 2023), or dialogue-optimized models like InstructGPT (Ouyang et al., 2022), ChatGPT (OpenAI, 2022), characterized by their extensive knowledge and language proficiency, have gained attention in many natural language processing (NLP) and natural language generation (NLG) tasks (Xi et al., 2023; Liu et al., 2023). LLMs possess the capability to process vast volumes of text, allowing for nuanced and context-aware understanding — a feature fundamental to aligning their output to match the user requests (Wang et al., 2023). The adaptability and versatility inherent in these models enable the development of tailored educational tools and applications, empowering students by providing personalized content that aligns with their learning styles and preferences (Cen et al., 2023; Qu et al., 2022; Embarak, 2022; Maghsudi et al., 2021b). By harnessing the capabilities of LLMs, educational platforms can offer interactive and adaptive content, contributing to enhanced engagement, deeper comprehension, and a more impactful learning experience for students across diverse disciplines.

We present a personalized summarization tool catered explicitly to the needs of humanities students, harnessing the capabilities of LLMs to provide tailored lecture summaries. LLM-based ap-

Proceedings of the 1st Workshop on Personalization of Generative AI Systems (PERSONALIZE 2024), pages 36–70
March 22, 2024

proaches to automatic summarization have shown promising results in capturing long-range dependencies, handling complex sentence structures, and producing coherent and contextually appropriate summaries (Bražinskas et al., 2020; Fabbri et al., 2020; Adams et al., 2022).

Our work addresses the following questions:

- RQ 1: How can we assess the diverse learning preferences of students for the development of a personalized summarization tool?

- RQ 2: How will users perceive and interact with this personalized summarization tool?

- RQ 3: How can we create a personalized tool without necessitating the acquisition of user data?

Section 3 presents the collection and preprocessing of transcribed university lectures on humanities studies, which were used as material for the summarizations. Section 4 presents the design of a survey, centered on scenarios and multiple options, followed by an analysis of survey results highlighting user trends and preferences. These insights have provided the requirements for the development of the summarization tool. Section 5 reports on the development of the tool, which was based on OpenAI's GPT-4 model (OpenAI, 2023), illustrating the integration of user preferences derived from the survey into the tool's framework. Finally, Section 6 presents an in-depth analysis of user engagement and perceptions regarding the tool, evaluating user feedback and satisfaction to assess the effectiveness and quality of personalized summaries.

2 Background and Related Work

2.1 Personalization and Education

In the field of education, personalization operates as a method rooted in evidence, centering on individualized learning abilities and study goals to tailor educational content to individual needs. This strategy entails adjusting, realigning, or reshaping components of the curriculum to harmonize with distinct individual requirements. As a result, it has the potential to tackle learning challenges faced by marginalized learners as well as to boost the overall efficacy of teaching and learning methodologies (Bhutoria, 2022; Yonezawa et al., 2012). Personalization stands as a fundamental principle in education, involving customized educational materials and accurate interventions to address distinct learning requirements. This approach, rooted in problem analysis and precise identification of student needs, constitutes the fundamental essence of personalized education (Cook et al., 2018; Fryer et al., 2017).

The emergence of Artificial Intelligence (AI) in education has broadened the scope for personalization, as AI-powered systems, such as Interactive Personalized Learning Spaces (PLS) and Intelligent Tutoring Systems (ITS), actively recommend tailored learning paths, customize educational content, and enrich learning experiences. These AI-driven platforms can effectively support pedagogical planning and language studies, as validated by empirical studies conducted across diverse educational environments (Qu et al., 2022; Zhang and Aslan, 2021; Hwang et al., 2020; Liu et al., 2017).

2.2 Personalizing Summaries for Educational Purposes

Summarization is an important NLP task aimed at condensing extensive amounts of information into cohesive and concise summaries. Its application spans diverse domains, from news articles to scientific papers, or legal documents (Altmami and Menai, 2022; El-Kassas et al., 2021; Kanapala et al., 2019). The surge in digital content since the 1950s has brought attention to automated summarization techniques, owing to the escalating need for efficient information retrieval and assimilation (Luhn, 1958; Allahyari et al., 2017).

The summarization of educational content has gained some attention even before the pandemic, due to the increasing popularity of E-learning. Yang et al. (2013) highlighted the shift toward mobile-based learning and the challenges posed by lengthy texts on such devices. Their research emphasized that well-crafted summaries could enhance learning experiences, particularly when aligning content with the device's constraints. Miller (2019) employed a BERT model to create a lecture summarization service for student use, indicating the move from traditional approaches to more AI-driven approaches.

On the other hand, personalization in summarization can be viewed as stylization, the ability to generate summaries that exhibit specific writing styles, tones, or levels of formality. By incorporating stylization techniques into the summarization

Department	Courses per Department	Transcripts per Department	Avg. Tokens per Transcript	Tot. Tokens	RTTR	MTLD
English	5	152	30843.8	819678.0	24.0	59.2
History	7	149	37662.3	604999.0	28.9	67.2
Philosophy	2	52	35588.4	498541.0	16.2	52.0
Psychology	2	43	51075.8	565734.0	21.7	63.2

Table 1: Descriptive statistics of the Yale Open Course lectures containing a total of 396 transcripts (RTTR = root type-token ratio; MTLD = measure of textual lexical diversity - both are average per transcript).

process, it becomes possible to align the summary's linguistic characteristics with the target audience (Díaz and Gervás, 2007; Yan et al., 2011; Móro et al., 2012). For example, in an educational setting, summaries can be stylized to match the level of understanding and familiarity of the students or their learning styles. This personalization allows students to engage with the summary in a manner that is comfortable and conducive to their learning style, enhancing their comprehension and retention of the lecture content.

While LLMs' capabilities have been investigated for providing personalized recommendations (Sidahmed et al., 2022; Chen, 2023; Lyu et al., 2023) and language learning interfaces (Kwon, 2023), they have not been studied for the specific task of lecture summarization. However, LLMs can generate summaries that are tailored to individual preferences and needs, and aligning the generated text to the user preference has been shown to improve usability by conditioning the model on specific prompts or incorporating external knowledge sources (Bai et al., 2022). LLMs can thus significantly contribute to the customization of lecture summaries.

3 Lecture Transcripts Acquisition and Preprocessing

We created a dataset comprising lecture transcripts suitable for students in the humanities, collecting data from Open Yale Courses (OYC)[1] (Kleiner, 2023). This platform hosts numerous courses across various disciplines, including subjects pertinent to humanities students such as English studies, history, psychology, and philosophy. Each course includes video content along with supplementary materials like syllabi, suggested readings, and searchable transcripts for every lecture, available in HTML format. Table 1 summarizes the dataset's descriptive statistics, including the average length of the transcripts. For tokenization, we utilized the English SpaCy model en_core_web_sm version 3.6.0 (Honnibal et al., 2020).[2]. We also assess lexical richness across topics by utilizing two metrics: the root type-token ratio (RTTR; Guiraud, 1958) and the textual lexical diversity measure (MTLD; McCarthy and Jarvis, 2010). These metrics were computed using the Lexical-Richness library (Shen, 2022), with MTLD employing a threshold of 0.72 to mitigate the influence of text length (Shen, 2022)[3]. The lectures demonstrate notably high scores for both RTTR and MTLD, highlighting a significant level of lexical diversity.

In the preprocessing phase, minimal adjustments were made to the data. Verbatim descriptions of non-verbal cues or actions, denoted as '[Crosstalk]', '[Laughter]', and '[Points at Student]', among others, which originated from the transcription of video elements, were removed. Additionally, indicators of pauses within the transcripts, delimited by timestamps such as [00:08:47], were also eliminated. These modifications were implemented to streamline the content for subsequent processing and analysis. Finally, incorporating the lecture transcripts into our summarization tool development involved storing essential details, including departmental information, course titles, professor names, and lecture session numbers, alongside the respective lecture transcripts.

4 Survey Design and Personalization Analysis

4.1 Scenario-Based Survey Development

To collect requirements for personalization options in lecture summarization, we conducted a scenario-based survey, designed to explore user preferences for a summarization tool. To replicate various decision-making scenarios for creating summaries, we crafted 36 scenarios.

For each survey question, we collected responses from 10 participants from the Amazon Mechanical Turk (AMT) platform, accumulating a total

[1]https://oyc.yale.edu/courses

[2]https://github.com/explosion/spacy-models/releases/tag/en_core_web_sm-3.3.0

[3]https://github.com/LSYS/LexicalRichness

of 360 responses for the entire survey. To ensure that our survey focused on a demographic close to university students, we restricted participants' age range to 20-25 and required a Turker approval rate exceeding 98%. To craft authentic and engaging scenarios, we asked the participants to adopt the persona of a humanities student, embedding academic requirements specific to this field within the instructions, for instance, that humanities students often engage in extensive reading, utilize critical and analytical thinking, and compose diverse types of essays (cf. Fig. 15).

Participants were presented with one scenario at a time, for example, *"You took a lot of courses this semester and you have a busy schedule balancing multiple courses and extracurricular activities. How would you prefer the summaries for your courses this semester?"*, and were then asked what type of summary would they need in this scenario, as multiple-choice questions. Scenarios provide realistic contexts for participants to envision practical usage and inform their preferences accordingly. The varied scenarios allow for a comprehensive exploration of diverse needs, aiding in tailoring the tool.

We identified different aspects of summary personalization and asked the participants to express their preference for each aspect by choosing one out of three given options. For each aspect, several scenarios were crafted to collect the requirements.

Length & Depth (8 scenarios): choice between

- Concise Summaries
- Moderately Detailed Summaries
- In-depth Summaries Summaries

here the participants could express their preference for brief, essential summaries to a more comprehensive coverage

Tone & Style (8 scenarios): choice between

- Formal and Academic Summaries
- Neutral and Objective Summaries
- Conversational and Informal Summaries

here the participants could request academic language or a casual, easy-to-understand conversational style

Complexity (7 scenarios): choice between

- Simplified Summaries
- Balanced or Neutral Summaries
- Detailed or Complex Summaries

here we catered to varying levels of complexity from straightforward to elaborate information

Summary Format (7 scenarios): choice between

- Bullet Point Summaries
- Paragraph-based Summaries
- Keyword Summaries

presenting information through bulleted lists, narrative-style elaboration, or structured keywords

Extra Features (6 scenarios): choice between

- Reflection Questions
- Entity Extraction
- Debate Focus

to enhance understanding by posing reflective queries, extracting essential entities, and highlighting debates or perspectives discussed in the lectures.

We also included a fourth option, "other", allowing participants to articulate any additional preferences for the summarization tool, in case the first three options did not match their preferences (cf. Figures (5-13) depict the scenarios presented to the participants and the responses, while Figures (14-15) showcase the assigned task in AMT platform).

4.2 Interpreting User Preferences and Identified Patterns

The analysis of the survey responses provides insights into the diverse preferences of students regarding the summarization of lecture content across various scenarios. For instance, in scenarios focusing on extracting specific information like names or crucial details from summaries, participants predominantly leaned towards the "Entity Extraction" option. This indicates a preference for tools that can efficiently identify and extract essential elements, aiding in information retrieval for specific queries. Similarly, scenarios involving the preparation of academic or formal materials, such as thesis content or consultation with professors, exhibited a strong inclination towards "Formal and Academic" style summaries, aligning with the need for structured and scholarly formats in such contexts.

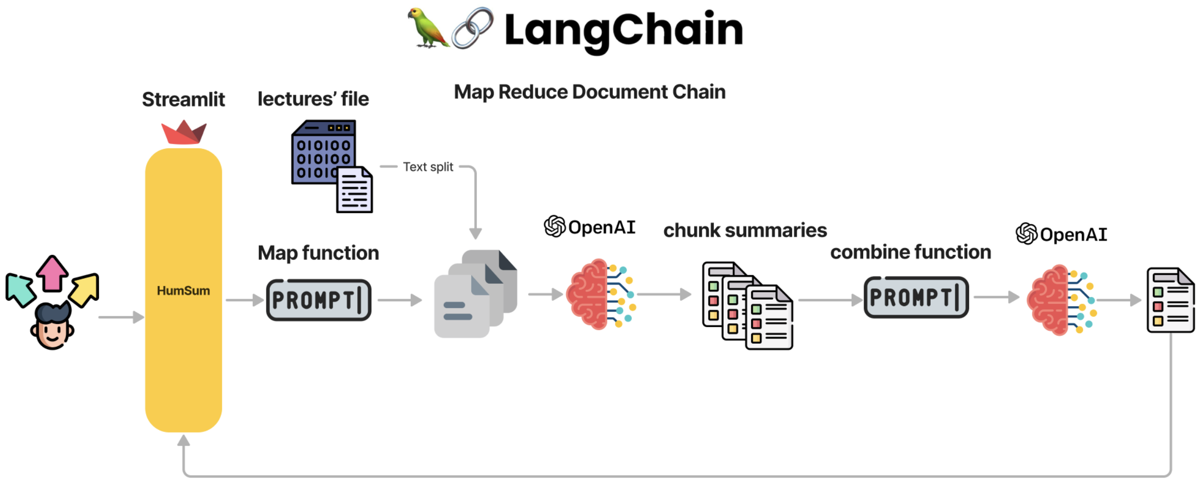

Figure 1: Summarization process using Streamlit, Langchain, and OpenAI's GPT4-8K model.

Furthermore, scenarios emphasizing comprehension and preparation for discussions or examinations revealed interesting trends. Participants favored different styles depending on their intended use. For instance, those preparing for discussions or debates leaned towards "Debate Focus" or "Reflection Question" options, indicating a preference for content stimulating critical thinking and discourse. In contrast, respondents preparing for exams or dealing with a busy schedule sought either "Concise Summaries" or "Moderately Detailed Summaries" that encapsulate essential concepts without excessive detail. Ultimately, regarding queries about the summary output style, participants tended to favor "Bullet point" or "Keyword" summaries when time constraints were a factor.

5 Summarization Tool Development

To craft the summarization tool, we leverage the insights collected from the analysis of scenario-driven survey outcomes. The tool aims at a straightforward design, enabling users to easily engage with the content, and personalizing the summaries to their needs.

5.1 User Interface

We created a user-friendly web-based tool named 'HumSum' designed specifically for the summarization process utilizing streamlit library[4]. HumSum features a convenient sidebar interface on the left-hand side, enabling users to input their preferences for the department, course, and session for which they seek a summary. Additionally, users can customize various aspects of the summary to suit their preferences as shown in figure 2. They can adjust the summary length, choosing from concise, moderate, or detailed options. Similarly, they can tailor the tone of the summary, opting for conversational, neutral, or formal language. Users can also select the complexity level — ranging from simplified to balanced to sophisticated — based on their comprehension needs. Moreover, the tool allows users to choose the format of the summary, whether it be in bullet-point, paragraph, or keyword style. Furthermore, users can opt for additional features such as extracting extra information like questions, entities, or debates from the lecture contents to enhance their summaries.

5.2 Langchain and LLM Chaining for Summarization

To generate lecture content summaries, we utilized Langchain[5], an open-source library compatible with Streamlit, offering access to various LLMs including those provided by OpenAI. Langchain encompasses diverse components such as Prompts, Memory, Chains, and Agents, which facilitate handling data stored in CSV files and support multiple NLP tasks involving information extraction and text generation. Specifically, we leveraged its 'Map Reduce Document Chain'[6] summarization chain feature, which allows for the summarization of long documents. As lecture contents are usually long and might exceed the input token limit of the LLM

[4]https://streamlit.io/

[5]https://www.langchain.com/

[6]https://js.langchain.com/docs/modules/chains/-popular/summarize

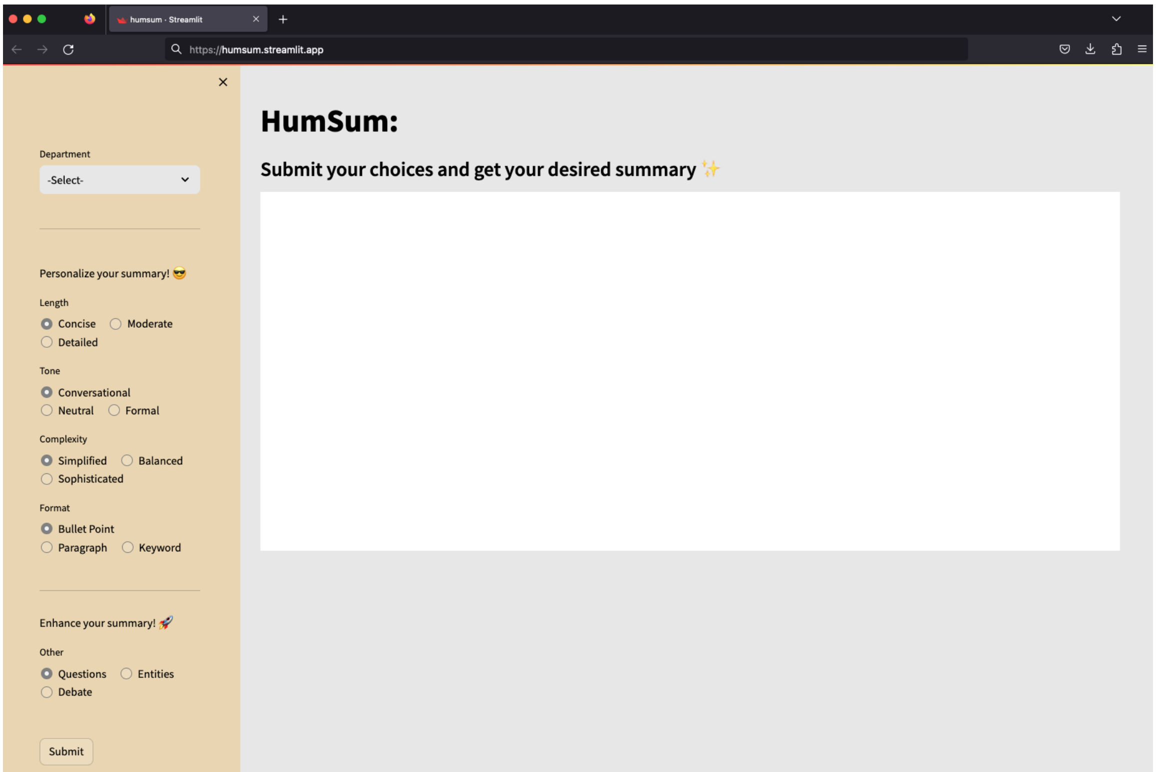

Figure 2: HumSum Interface - The slider on the left contains the customization options for the user.

models, we used this setup that enables the summarization of extensive document collections within a map-reduce style architecture. Langchain's Map Reduce involves dividing documents into smaller chunks that fit within the token limit of the model. For our lectures, we've specifically set this limit to 2000 tokens per chunk using Langchain's "Recursive Character Text Splitter"[7]. Each chunk is then summarized giving the user-selected preferences submitted via HumSum as instruction to the model (the map function), followed by the consolidation of these summaries into a final summary (combine or reduce function). To create the summaries, we made use of OpenAI's GPT-4-8k model (cf. Fig. 3 for the custom map prompt and Fig. 1 for the summarization process).

Summarise the lecture content from the course titled "{course_title}" for session {session_num} in the {department_name} department. The summary should meet the specified criteria:

- Length: {length_type}
- Tone: {tone_type}
- Complexity level: {complexity_type}
- Format: {format_type}
- Additional features: {other_features}

If the additional feature is "questions," generate five questions based on the lecture content. If it is "entities," extract names, dates, and locations from the lecture content. If the additional feature is "debate," extract opposing viewpoints or presented ideas from the lecture content.

Figure 3: Custom prompt to replace the default map prompt.

6 Assessing Preferences and Effectiveness through User Evaluation

6.1 Assessing User Preferences based on the Given Options

To evaluate the effectiveness of the HumSum summarization tool and the user-customizable options, we carried out a follow-up survey by recruiting three participants per scenario, all of whom were humanities students. This second survey utilized the same scenarios as the initial one, albeit with slight modifications, including additional information prompting participants to select their department, course title, and session. For instance the first survey's question such as *You took a lot of courses this semester and you have a busy schedule balancing multiple courses and extracurricular*

[7] https://python.langchain.com/docs/-modules/data_connection/document_transformers-/text_splitters/recursive_text_splitter

activities. how would you prefer the summaries for your courses this semester? has been modified to *Imagine yourself as an English student at Yale University. You would like to create a summary for Session Three of the course "Introduction to Theory of Literature", keeping in mind that you took a lot of courses this semester and you have a busy schedule balancing multiple courses and extracurricular activities.*. The students were asked to explore the tool and select their preferred options for personalizing the summaries of lecture transcripts. Throughout this process, no additional instructions were provided, and the students navigated the tool seamlessly without asking any questions or encountering difficulties. Once they finalized their choices, the students submitted them, and the HumSum tool displayed the resulting summarization based on their preferences (cf. Fig. 16 - 20 for a sample user journey).

The primary aims of this subsequent survey were twofold: firstly, to assess whether participants' choices remained consistent now that they had complete access to the application interface, and secondly, to discern the most prevalent user choices overall. This analysis aims to identify recurring preferences among participants, to inform future further development of the tool. Figure 4 shows the personalization options that received the highest number of selections per aspect (cf. Fig. 21 to 23 to see the frequency of all the provided options per aspect).

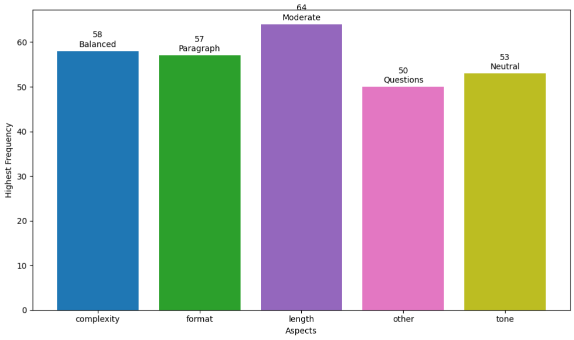

Figure 4: The options that received the highest number of selections within each aspect.

As depicted in Figure 4, users predominantly favored middle-ground choices as their primary selection, opting for moderate length, neutral tone, or balanced complexity. This tendency might be attributed to users' cautious approach, potentially stemming from their lack of prior experience with the tool and limited exposure to the generated summaries or real-world scenario analysis of the tool's efficacy.

This assertion finds additional support in Figure 24 in the appendix, displaying a comparison between our anticipated selections and the actual number of user preferences aligned with our expectations. Notably, options entailing more complex summaries, tailored to specific scenario requirements, were consistently not chosen. When questioned about this, users unanimously expressed concerns about handling such complexity, especially when required to memorize summaries for classroom discussions or exams.

While the infrequent selection of non-middle-ground options suggests a cautious approach by users, it doesn't imply that these options will never be chosen by potential users of the tool. We recognize the importance of further exploration and subsequent iterations to assess the tool's performance in actual classroom settings. Collecting real classroom data from multiple usage cycles by students in subsequent iterations will provide valuable insights into the tool's practical utility and user preferences.

6.2 Assessing the Tool Efficacy and the Personalization of the Summaries

To collect feedback regarding the effectiveness of the summarization tool and the resultant summaries based on users' selections, we administered a concise questionnaire featuring six rating-scale questions ranging from 1 to 5, as shown below. Overall, the average rating across all questions was 4 out of 5, indicating a favorable reception of this tool for summarizing humanities-related content, with users showing a high level of satisfaction when navigating the tool. The outcome of this questionnaire is displayed in Figure 25.

1. How satisfied are you with the summarization tool's user interface and ease of navigation?

2. Rate the accuracy of the summaries generated based on your selected preferences.

3. How well did the summaries reflect your specified preferences for length, tone, complexity, and format?

4. Rate the relevance and usefulness of the additional features (e.g., questions, entity extraction, debates) provided with the summaries.

5. How would you rate the personalization options provided in this summarization tool?

6. How likely are you to recommend this summarization tool to other students?

7 Conclusion and Future Work

In this study, we developed a personalized summarization tool, "HumSum," tailored for humanities students to efficiently condense lecture transcripts. We carried out a scenario-driven survey to identify and integrate personalization features within the tool, accommodating users' preferences for summary types. Utilizing the Streamlit library, Langchain, and OpenAI's GPT-4-8k model facilitated the tool's development and the summarization process. Subsequently, we evaluated the tool's effectiveness through a user journey based on the pre-defined scenarios and a straightforward rating questionnaire, to evaluate the tool's performance along with the users' anticipated outcomes.

This tool offers some advantages: it offers personalized summarization without relying on users' private information, in addition to being an easy-to-use tool. However, for future improvements, we aim to implement a feedback integration mechanism, enabling users to provide real-time insights to refine the summary output iteratively. Addressing the issue of hallucination, a critical concern in utilizing language models remains a future task (Huang et al., 2023). Additionally, integrating external knowledge sources, such as highlighting key information or providing links to supplementary materials like Wikipedia, could significantly enrich the tool's usability and enhance the learning experience for students.

Limitations

While our research yielded valuable insights, addressing its limitations is crucial. One such constraint, as identified in Section 6.1, pertains to users' predominant inclination towards moderate options within the summarization tool. This predilection could be attributed to users' unfamiliarity with the tool, limiting their perception of how specific options might shape a customized summary. Additionally, the scenarios employed might not have effectively elicited all feasible options, suggesting that continuous iterations over time could offer a more comprehensive understanding of the tool's efficacy and available options.

Regarding the Map Reduce method employed, its merits lie in scalability for larger documents and the independent nature of LLM calls on individual document segments, enabling concurrent processing. Nonetheless, several drawbacks should be considered. Primarily, this approach requires higher LLM calls to create a summary for one large document. Furthermore, there is a risk of information loss during the final reduction or combination phase. These limitations underscore areas for potential refinement and call for cautious consideration when employing the method in practical applications.

Ethics Statement

The turkers we recruited on the AMT platform as well as the students maintain their anonymity, a practice aligned with ethical norms within the community. They were recruited voluntarily and provided a written consent form to participate in the study and were allowed to opt out at any point in time. Moreover, the AMT workers were compensated following the norms and regulations of the AMT platform for their time and effort spent on our tasks. We encouraged feedback and offered to promptly address any concerns or issues that might arise during the research process. However, we did not record any issues and we received positive feedback regarding the experiments.

The icons showcased in Figure 1 have been sourced from the freely available collection at https://www.flaticon.com/.

Acknowledgements

We extend our sincere thanks to the AMT participants and the students whose invaluable contributions were essential in completing this research.

References

David Adams, Gandharv Suri, and Yllias Chali. 2022. Combining state-of-the-art models with maximal marginal relevance for few-shot and zero-shot multi-document summarization. *arXiv preprint arXiv:2211.10808*.

Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assefi, Saeid Safaei, Elizabeth D Trippe, Juan B Gutierrez, and Krys Kochut. 2017. Text summarization techniques: a brief survey. *arXiv preprint arXiv:1707.02268*.

Nouf Ibrahim Altmami and Mohamed El Bachir Menai. 2022. Automatic summarization of scientific articles:

A survey. *Journal of King Saud University-Computer and Information Sciences*, 34(4):1011–1028.

Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. 2022. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*.

Aditi Bhutoria. 2022. Personalized education and artificial intelligence in the united states, china, and india: A systematic review using a human-in-the-loop model. *Computers and Education: Artificial Intelligence*, 3:100068.

Arthur Bražinskas, Mirella Lapata, and Ivan Titov. 2020. Few-shot learning for opinion summarization. *arXiv preprint arXiv:2004.14884*.

Chenghong Cen, Guang Luo, Lujia Li, Yilin Liang, Kang Li, Tan Jiang, and Qiang Xiong. 2023. User-centered software design: User interface redesign for blockly–electron, artificial intelligence educational software for primary and secondary schools. *Sustainability*, 15(6):5232.

Zheng Chen. 2023. Palr: Personalization aware llms for recommendation. *arXiv preprint arXiv:2305.07622*.

Clayton R. Cook, Stephen P. Kilgus, and Matthew K. Burns. 2018. Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology*, 66:4–10. Advancing the Science and Practice of Precision Education.

Alberto Díaz and Pablo Gervás. 2007. User-model based personalized summarization. *Information Processing & Management*, 43(6):1715–1734.

Wafaa S El-Kassas, Cherif R Salama, Ahmed A Rafea, and Hoda K Mohamed. 2021. Automatic text summarization: A comprehensive survey. *Expert systems with applications*, 165:113679.

Ossama H Embarak. 2022. Internet of behaviour (iob)-based ai models for personalized smart education systems. *Procedia Computer Science*, 203:103–110.

Alexander R Fabbri, Simeng Han, Haoyuan Li, Haoran Li, Marjan Ghazvininejad, Shafiq Joty, Dragomir Radev, and Yashar Mehdad. 2020. Improving zero and few-shot abstractive summarization with intermediate fine-tuning and data augmentation. *arXiv preprint arXiv:2010.12836*.

Muhammad Ashraf Fauzi. 2022. E-learning in higher education institutions during covid-19 pandemic: current and future trends through bibliometric analysis. *Heliyon*.

Luke K. Fryer, Mary Ainley, Andrew Thompson, Aaron Gibson, and Zelinda Sherlock. 2017. Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners. *Computers in Human Behavior*, 75:461–468.

P. Guiraud. 1958. *Problèmes et Méthodes de La Statistique Linguistique*. Dodrecht: D. Reidel.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python.

Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. 2023. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions.

Gwo-Jen Hwang, Han-Yu Sung, Shao-Chen Chang, and Xing-Ci Huang. 2020. A fuzzy expert system-based adaptive learning approach to improving students' learning performances by considering affective and cognitive factors. *Computers and Education: Artificial Intelligence*, 1:100003.

Ambedkar Kanapala, Sukomal Pal, and Rajendra Pamula. 2019. Text summarization from legal documents: a survey. *Artificial Intelligence Review*, 51:371–402.

Diana E. E. Kleiner. 2023. Open yale courses (oyc).

Anis Koubaa. 2023. Gpt-4 vs. gpt-3.5: A concise showdown.

Taeahn Kwon. 2023. *Interfaces for Personalized Language Learning with Generative Language Models*. Ph.D. thesis, Columbia University.

Danny Yen-Ting Liu, Kathryn Bartimote-Aufflick, Abelardo Pardo, and Adam J. Bridgeman. 2017. Data-driven personalization of student learning support in higher education. In Alejandro Peña-Ayala, editor, *Learning Analytics: Fundaments, Applications, and Trends: A View of the Current State of the Art to Enhance e-Learning*, pages 143–169. Springer International Publishing, Cham.

Yiheng Liu, Tianle Han, Siyuan Ma, Jiayue Zhang, Yuanyuan Yang, Jiaming Tian, Hao He, Antong Li, Mengshen He, Zhengliang Liu, et al. 2023. Summary of chatgpt-related research and perspective towards the future of large language models. *Meta-Radiology*, page 100017.

Hans Peter Luhn. 1958. The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2):159–165.

Maggie Lynch. 2020. E-learning during a global pandemic. *Asian Journal of Distance Education*, 15(1):189–195.

Hanjia Lyu, Song Jiang, Hanqing Zeng, Qifan Wang, Si Zhang, Ren Chen, Chris Leung, Jiajie Tang, Yinglong Xia, and Jiebo Luo. 2023. Llm-rec: Personalized recommendation via prompting large language models.

Abdelsalam M Maatuk, Ebitisam K Elberkawi, Shadi Aljawarneh, Hasan Rashaideh, and Hadeel Alharbi. 2022. The covid-19 pandemic and e-learning: challenges and opportunities from the perspective of students and instructors. *Journal of computing in higher education*, 34(1):21–38.

Setareh Maghsudi, Andrew Lan, Jie Xu, and Mihaela van Der Schaar. 2021a. Personalized education in the artificial intelligence era: what to expect next. *IEEE Signal Processing Magazine*, 38(3):37–50.

Setareh Maghsudi, Andrew Lan, Jie Xu, and Mihaela van der Schaar. 2021b. Personalized education in the artificial intelligence era: What to expect next. *IEEE Signal Processing Magazine*, 38(3):37–50.

Philip M McCarthy and Scott Jarvis. 2010. Mtld, vocd-d, and hd-d: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior research methods*, 42(2):381–392.

Derek Miller. 2019. Leveraging bert for extractive text summarization on lectures. *arXiv preprint arXiv:1906.04165*.

Róbert Móro et al. 2012. Personalized text summarization based on important terms identification. In *2012 23rd International Workshop on Database and Expert Systems Applications*, pages 131–135. IEEE.

OpenAI. 2022. OpenAI: Introducing ChatGPT. [Online; posted 30-November-2022].

OpenAI. 2023. Gpt-4 technical report.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback.

Jianjing Qu, Yanan Zhao, and Yongping Xie. 2022. Artificial intelligence leads the reform of education models. *Systems Research and Behavioral Science*, 39(3):581–588.

Rajapandian Radha, K Mahalakshmi, V Sathish Kumar, and AR Saravanakumar. 2020. E-learning during lockdown of covid-19 pandemic: A global perspective. *International journal of control and automation*, 13(4):1088–1099.

Rolf Reber, Elizabeth A Canning, and Judith M Harackiewicz. 2018. Personalized education to increase interest. *Current directions in psychological science*, 27(6):449–454.

Victoriya V Sadovaya, Olga V Korshunova, and Zhumagali Zh Nauruzbay. 2016. Personalized education strategies. *International Electronic Journal of Mathematics Education*, 11(1):199–209.

Lucas Shen. 2022. LexicalRichness: A small module to compute textual lexical richness.

Hakim Sidahmed, Sameer Ahuja, and Manoj Kumar Tiwari. 2022. Generating personalized recommendations via large language models (llms).

Leonard Tetzlaff, Florian Schmiedek, and Garvin Brod. 2021. Developing personalized education: A dynamic framework. *Educational Psychology Review*, 33:863–882.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models.

Yufei Wang, Wanjun Zhong, Liangyou Li, Fei Mi, Xingshan Zeng, Wenyong Huang, Lifeng Shang, Xin Jiang, and Qun Liu. 2023. Aligning large language models with human: A survey. *arXiv preprint arXiv:2307.12966*.

Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, et al. 2023. The rise and potential of large language model based agents: A survey. *arXiv preprint arXiv:2309.07864*.

Rui Yan, Jian-Yun Nie, and Xiaoming Li. 2011. Summarize what you are interested in: An optimization framework for interactive personalized summarization. In *Proceedings of the 2011 conference on empirical methods in natural language processing*, pages 1342–1351.

Guangbing Yang, Nian-Shing Chen, Kinshuk, Erkki Sutinen, Terry Anderson, and Dunwei Wen. 2013. The effectiveness of automatic text summarization in mobile learning contexts. *Computers & Education*, 68:233–243.

Susan Yonezawa, Larry McClure, and Makeba Jones. 2012. Personalization in schools. *Education Digest*, 78(2):41–47.

Ke Zhang and Ayse Begum Aslan. 2021. Ai technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, 2:100025.

A Appendix

A.1 Comparison of Expected and Actual Responses: Analysis of AMT Participants' Survey Data

The displayed bar plots represent the responses obtained from AMT participants for each scenario. The orange color indicates the anticipated or expected responses from participants. In the majority of cases, the participants' responses aligned with our anticipated patterns.

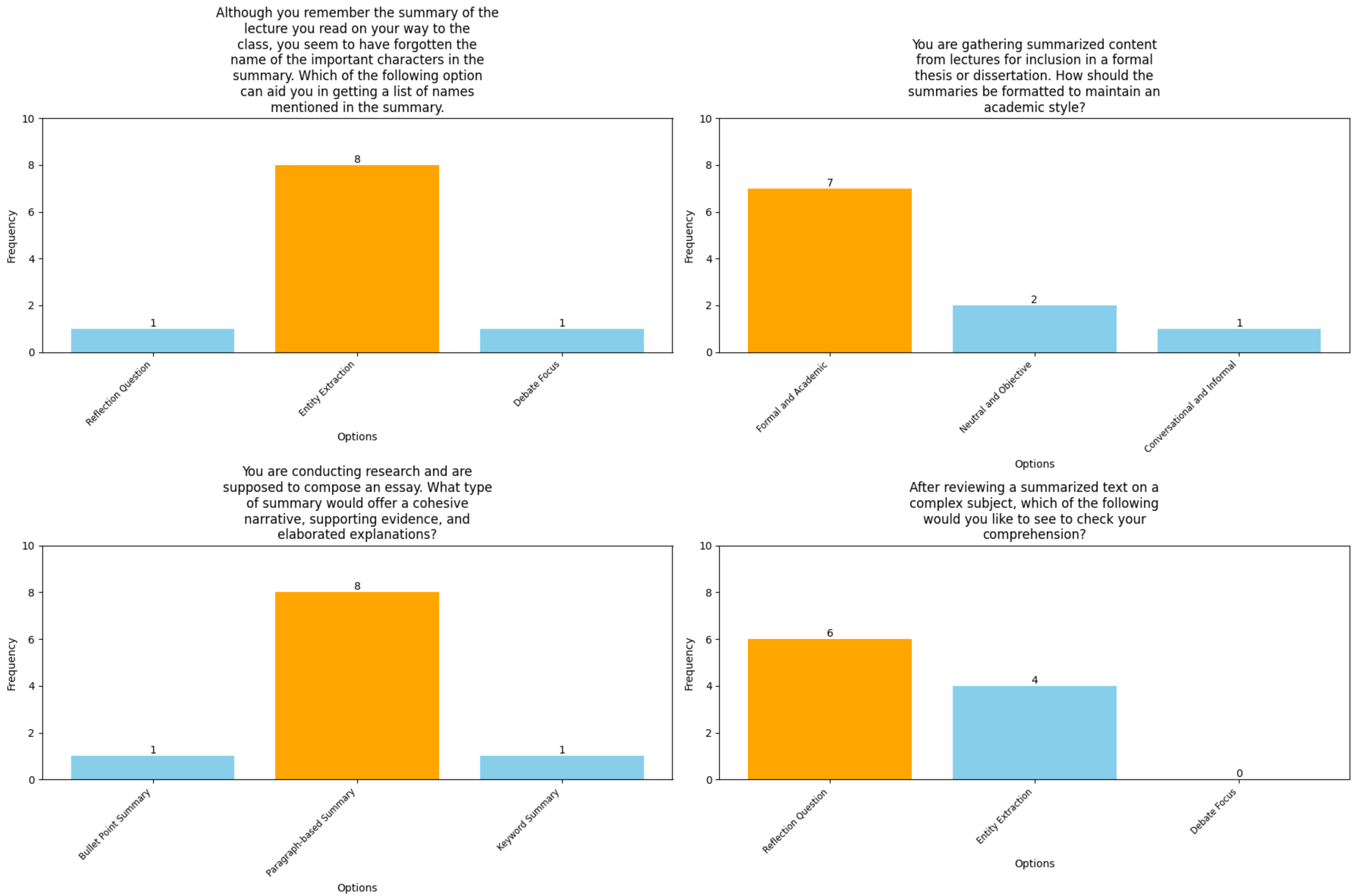

Figure 5: Results from scenarios 1-4

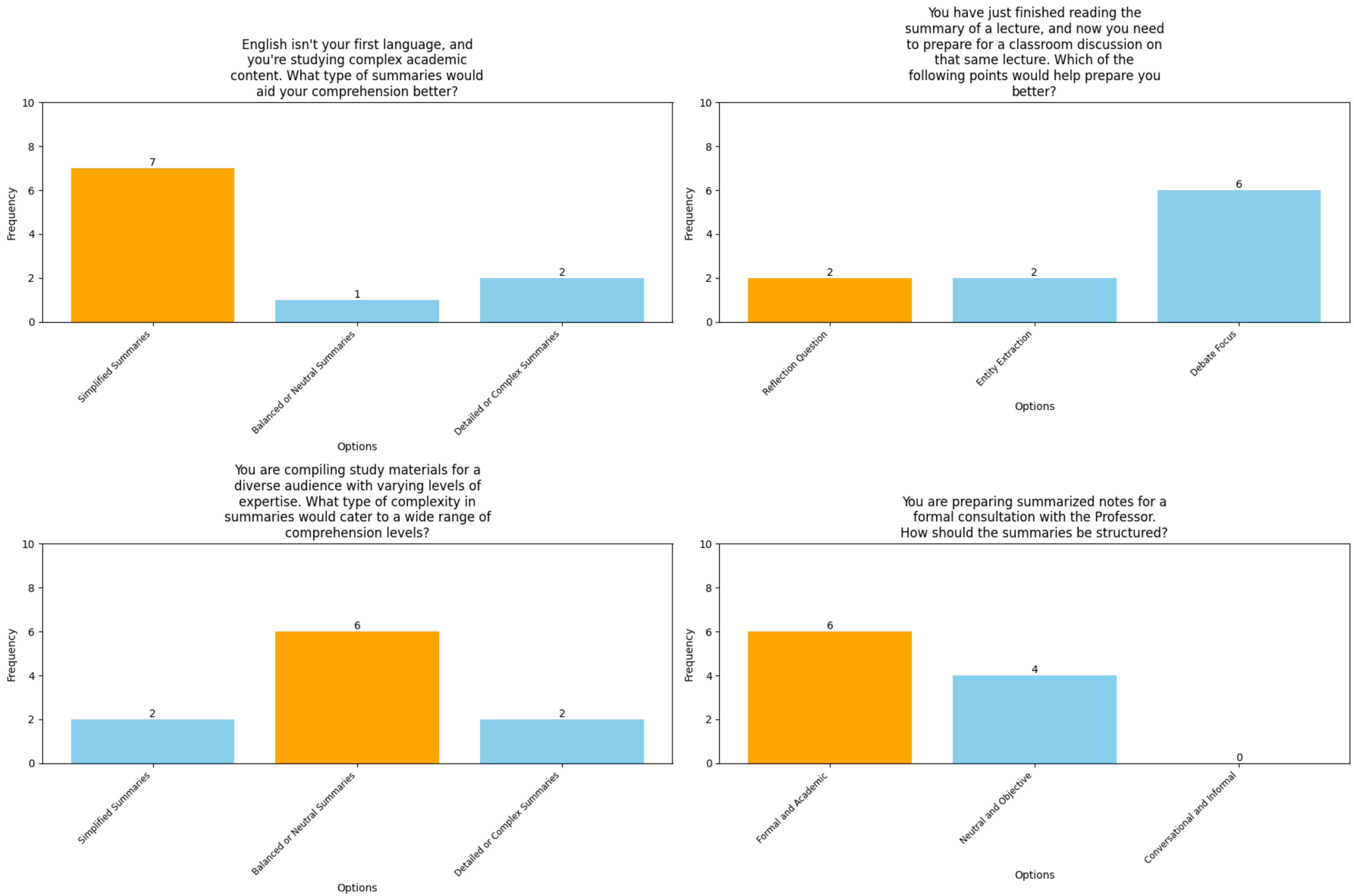

Figure 6: Results from scenarios 5-8

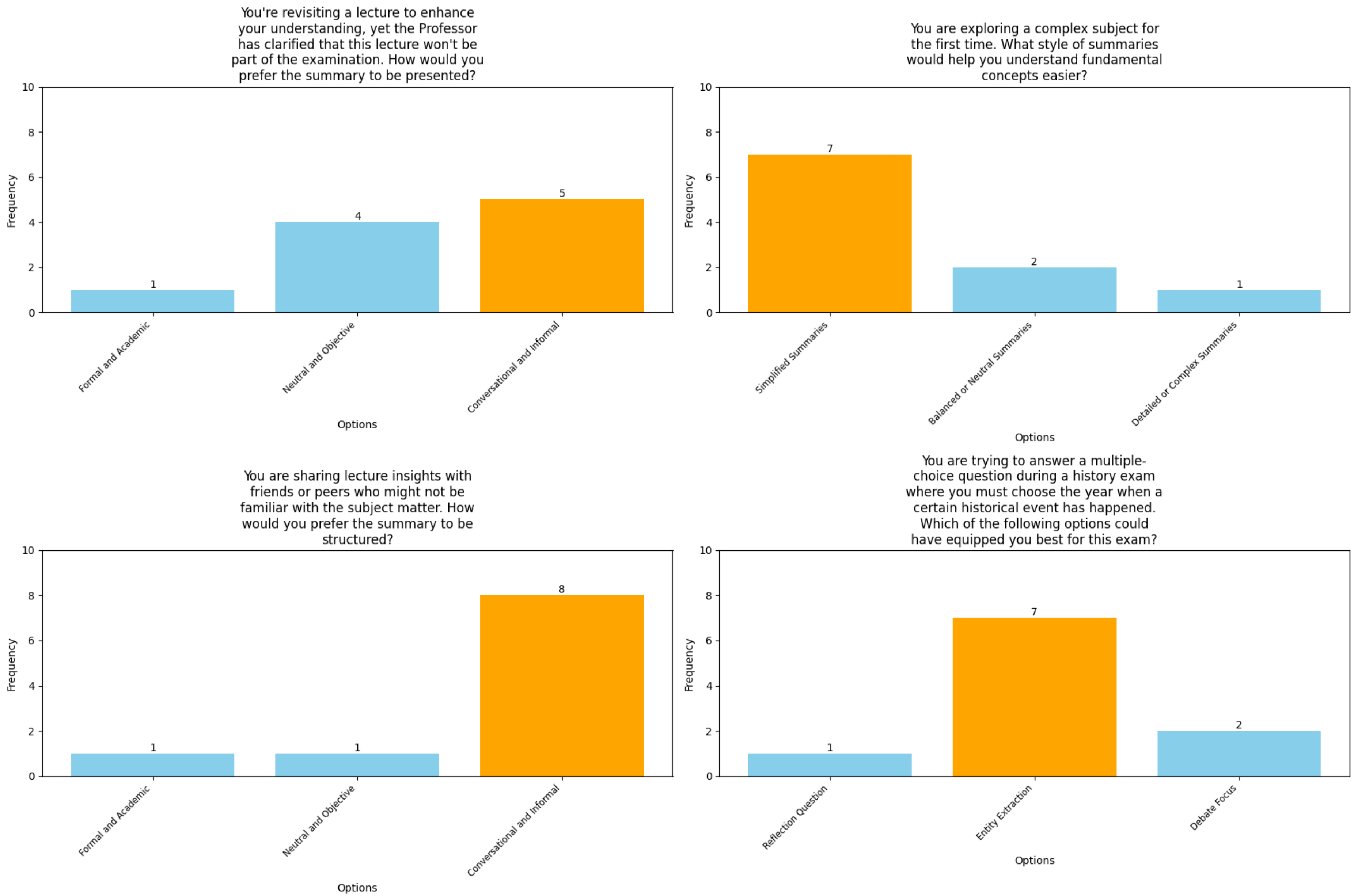

Figure 7: Results from scenarios 9-12

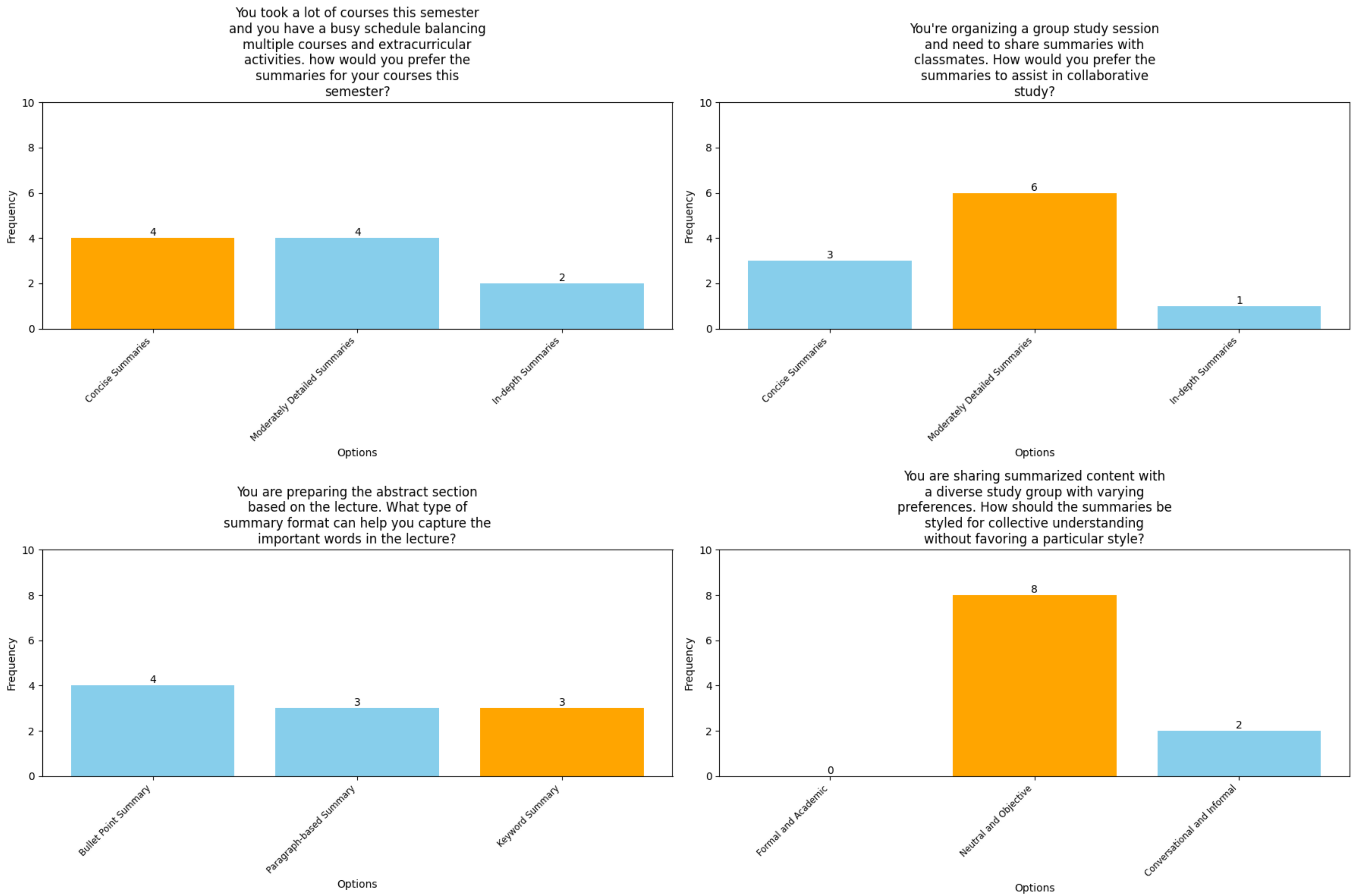

Figure 8: Results from scenarios 13-16

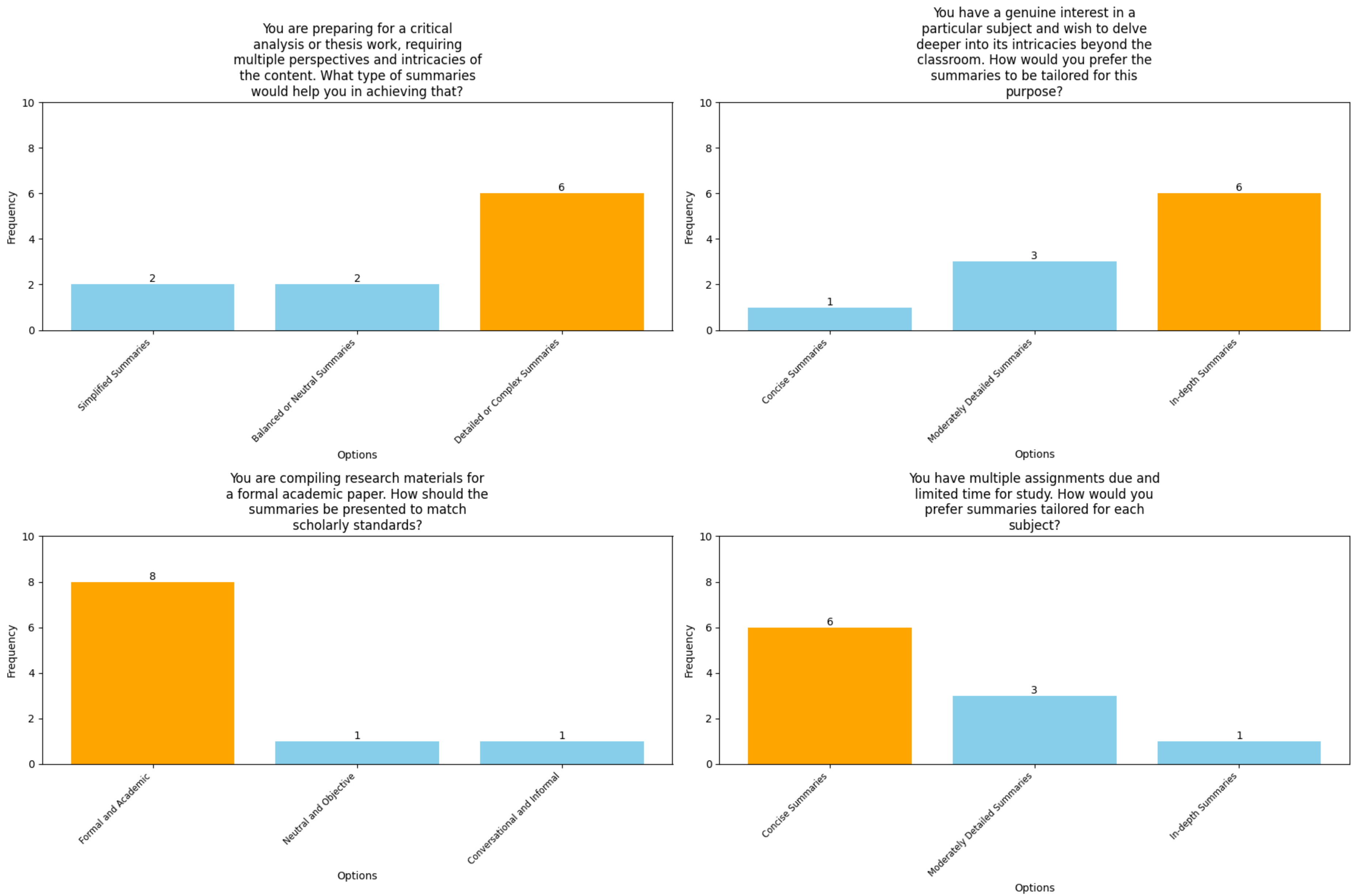

Figure 9: Results from scenarios 17-20

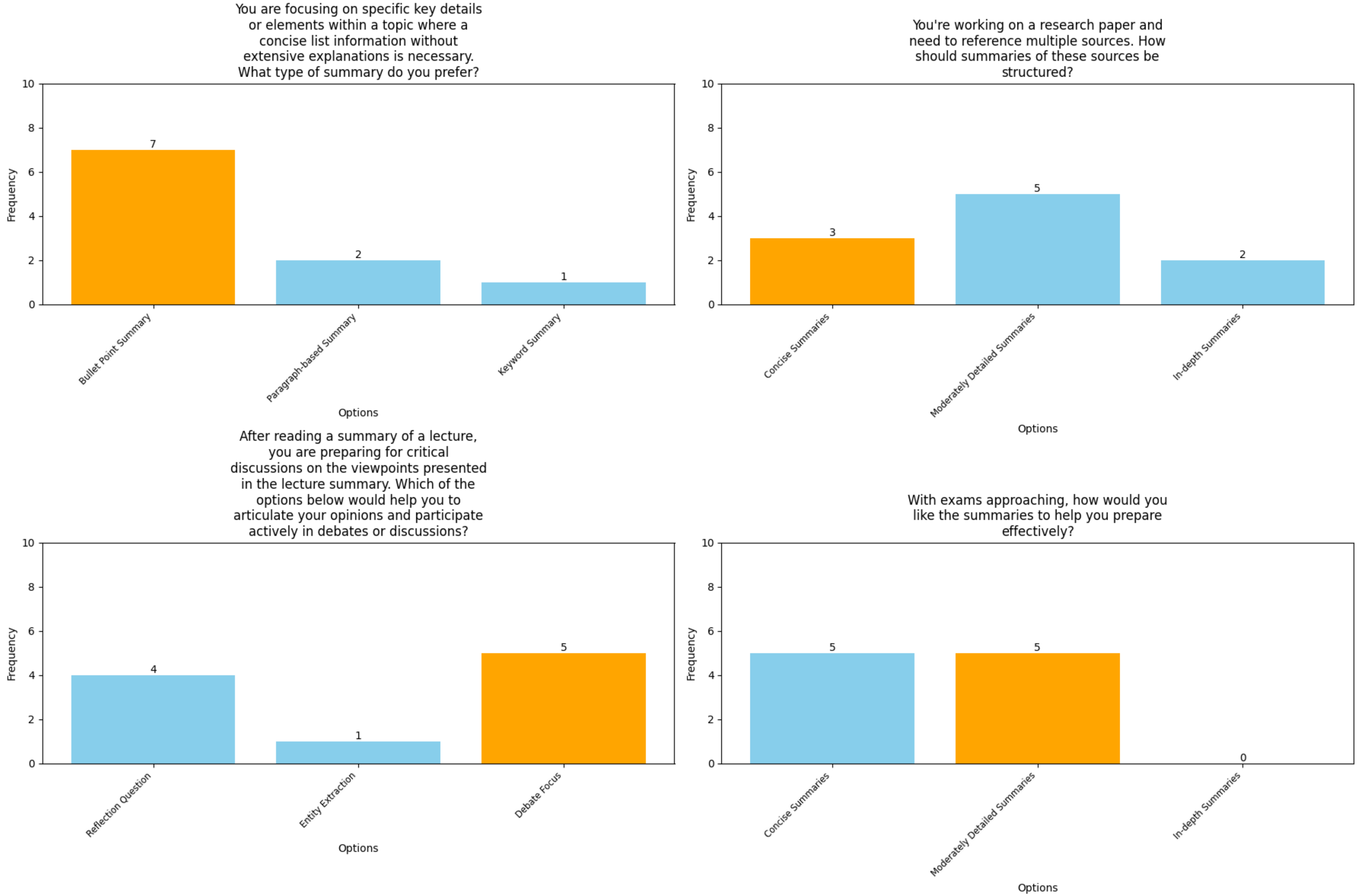

Figure 10: Results from scenarios 21-24

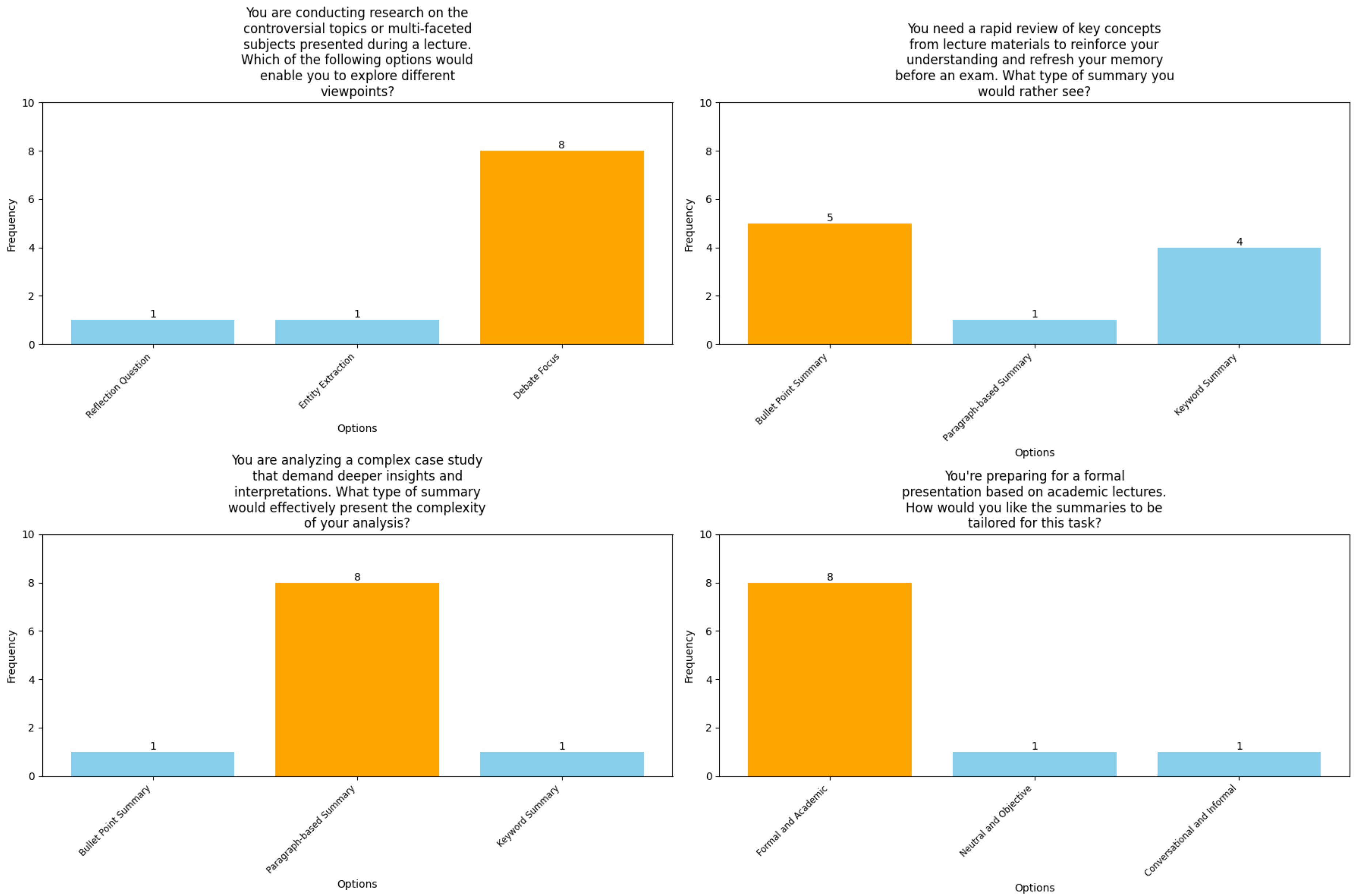

Figure 11: Results from scenarios 25-28

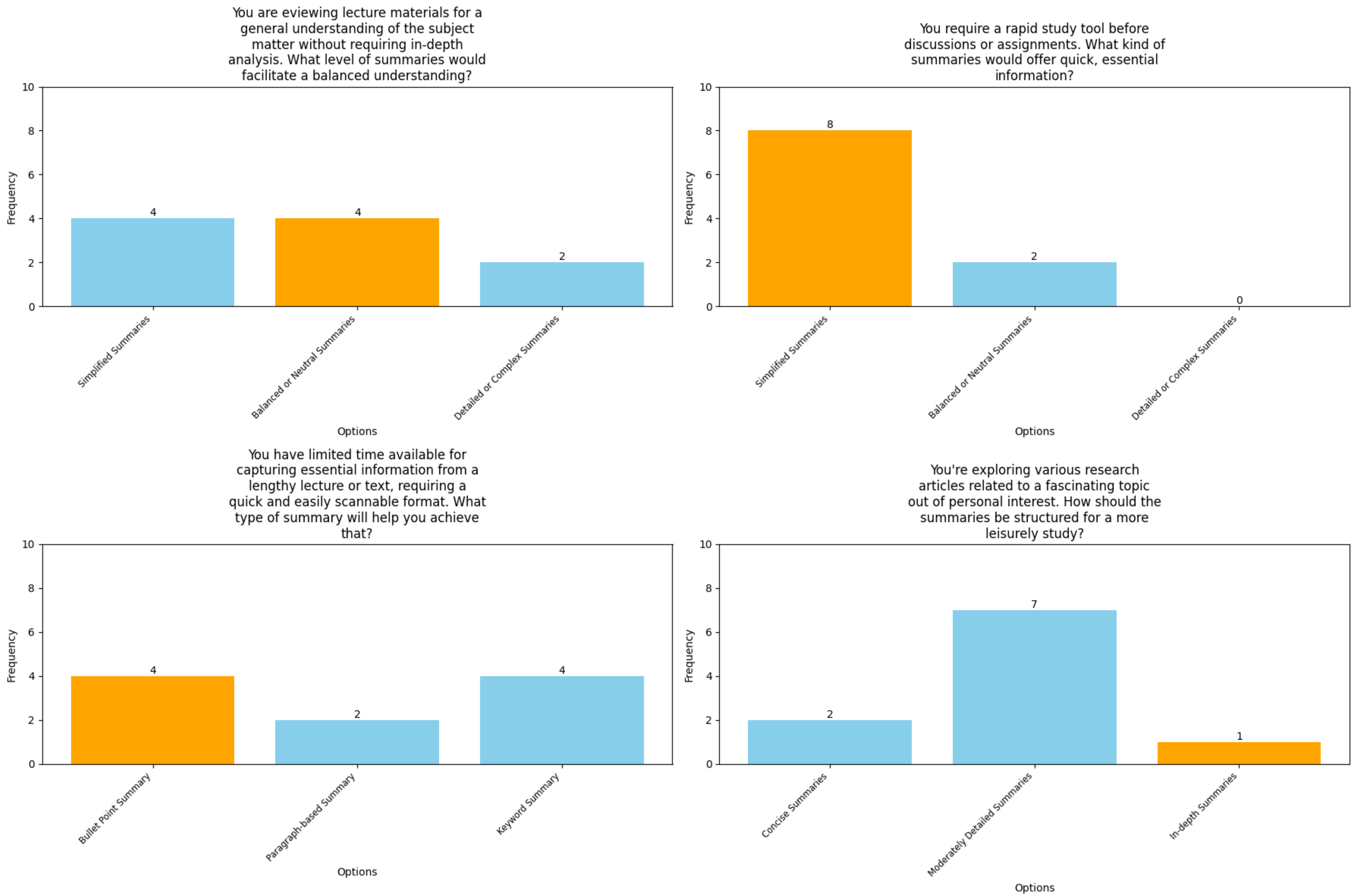

Figure 12: Results from scenarios 29-32

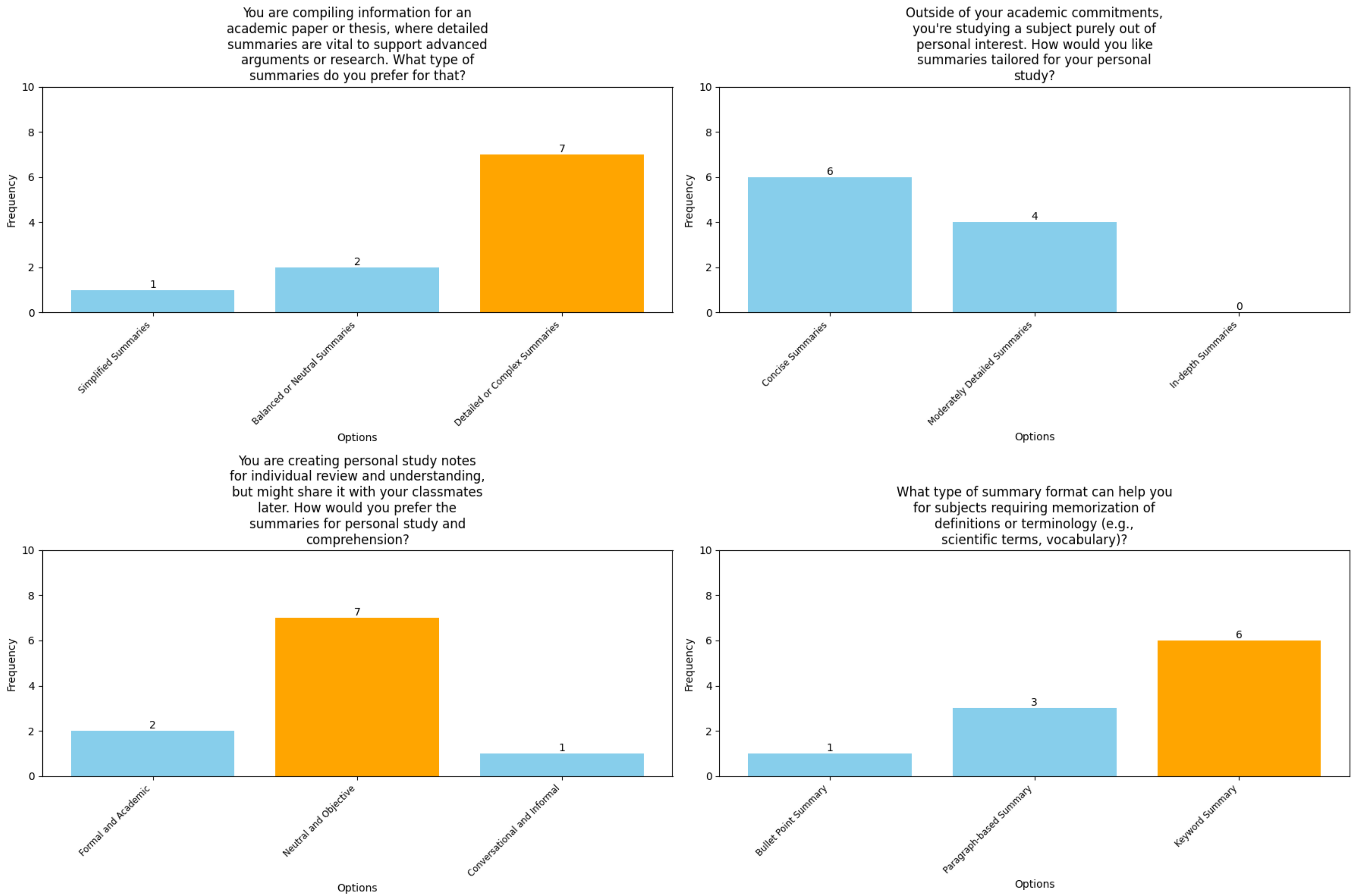

Figure 13: Results from scenarios 33-36

A.2 Example Scenario and Instructions Presented to AMT Participants

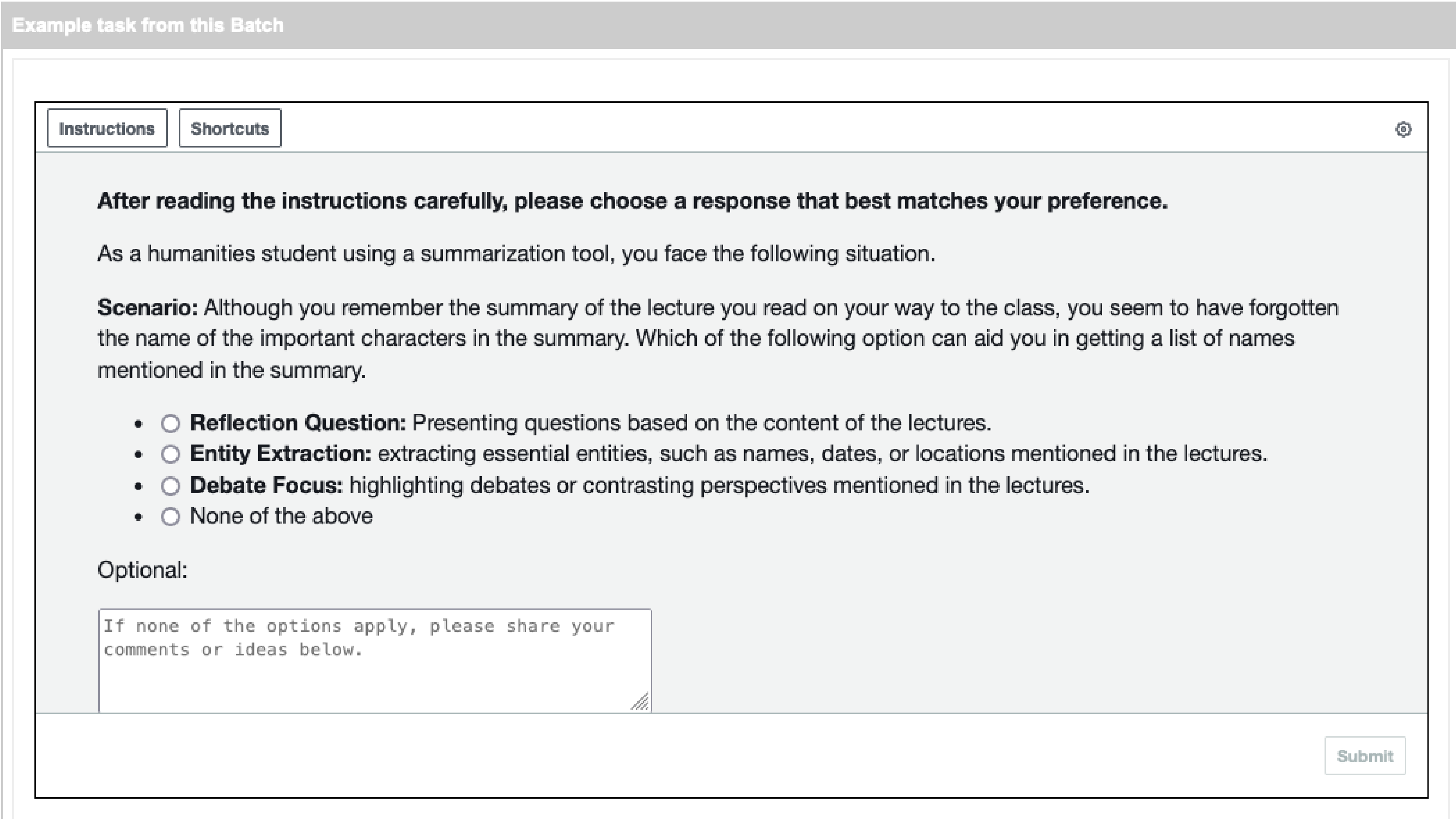

Figure 14: Example Scenario and options presented to AMT participants.

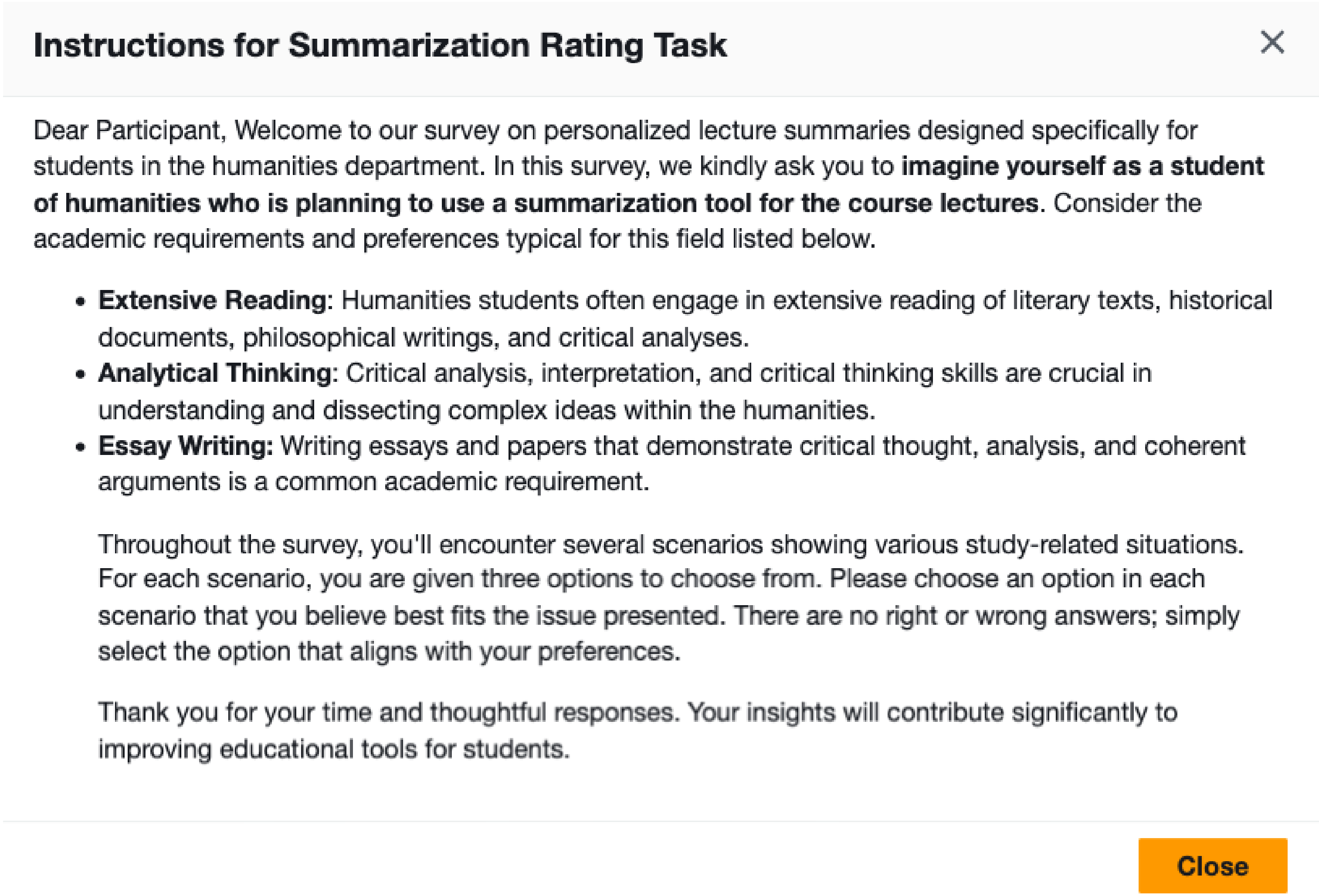

Figure 15: Instruction presented to AMT participants.

A.3 Sample User Journey to create Personalized Summary with HumSum

This user journey presented here is formulated following a modified version of the initial survey's question below.

Initial survey's question: You took a lot of courses this semester and you have a busy schedule balancing multiple courses and extracurricular activities. how would you prefer the summaries for your courses this semester?

Modified question for this user journey: Imagine yourself as an English student at Yale University. You would like to create a summary for Session Three of the course "Introduction to Theory of Literature", keeping in mind that you took a lot of courses this semester and you have a busy schedule balancing multiple courses and extracurricular activities.

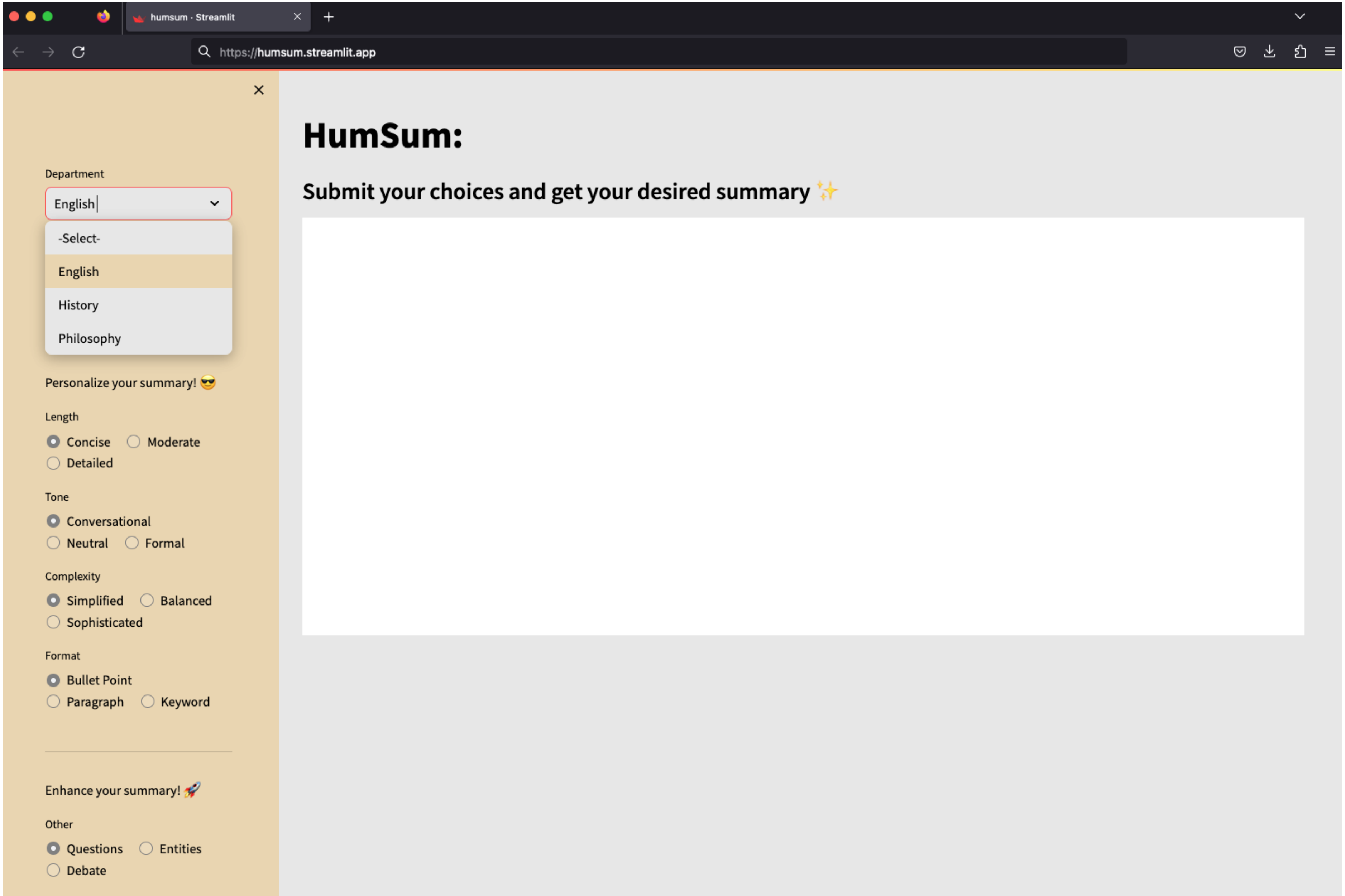

Figure 16: User can select among the humanities departments. The provided options to choose from are "English", "History", "Philosophy", and "Psychology" based on the available departments extracted from OpenYale courses described in section 3.

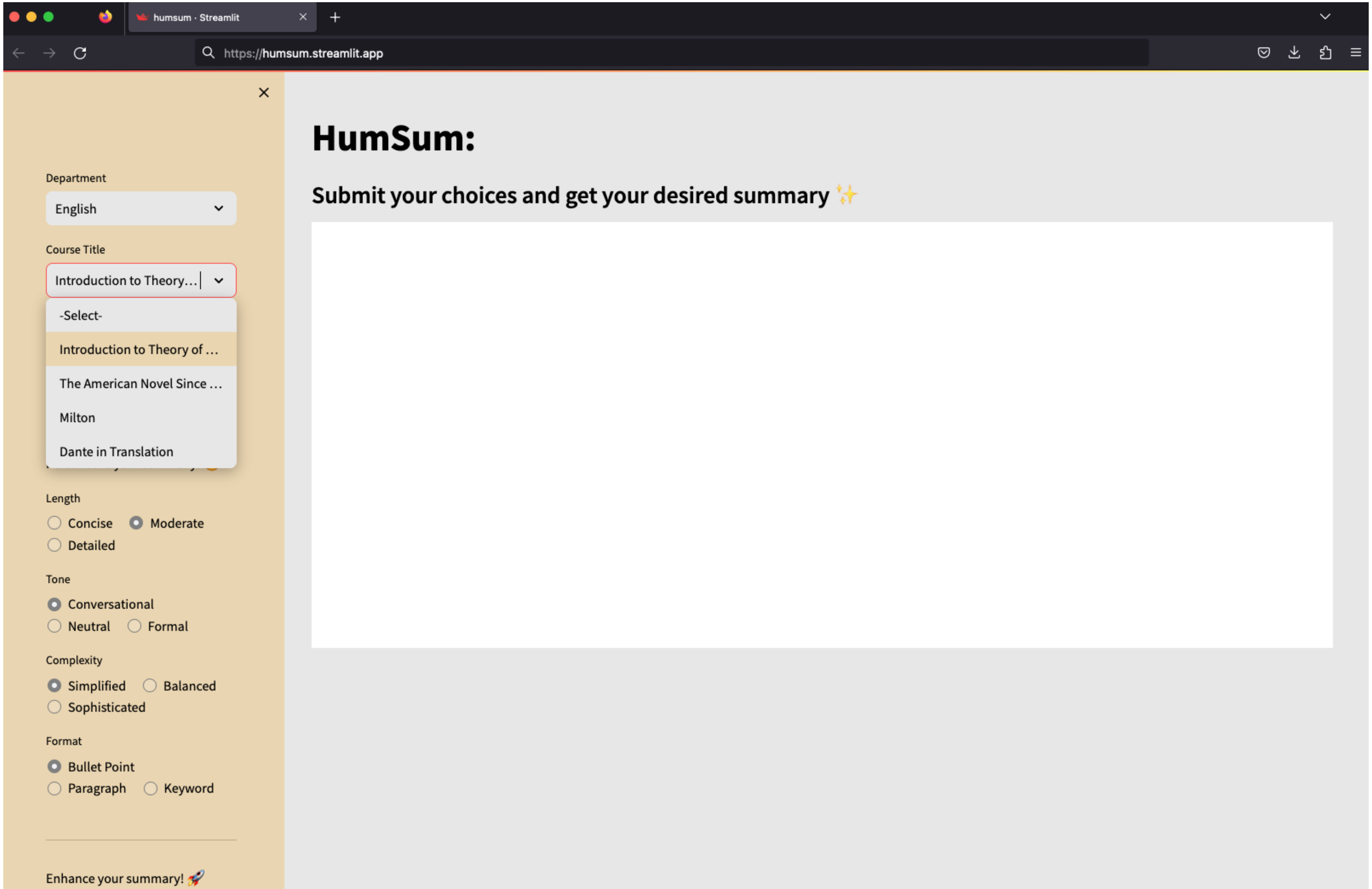

Figure 17: User can select the course title based on the courses extracted from OpenYale courses described in section 3.

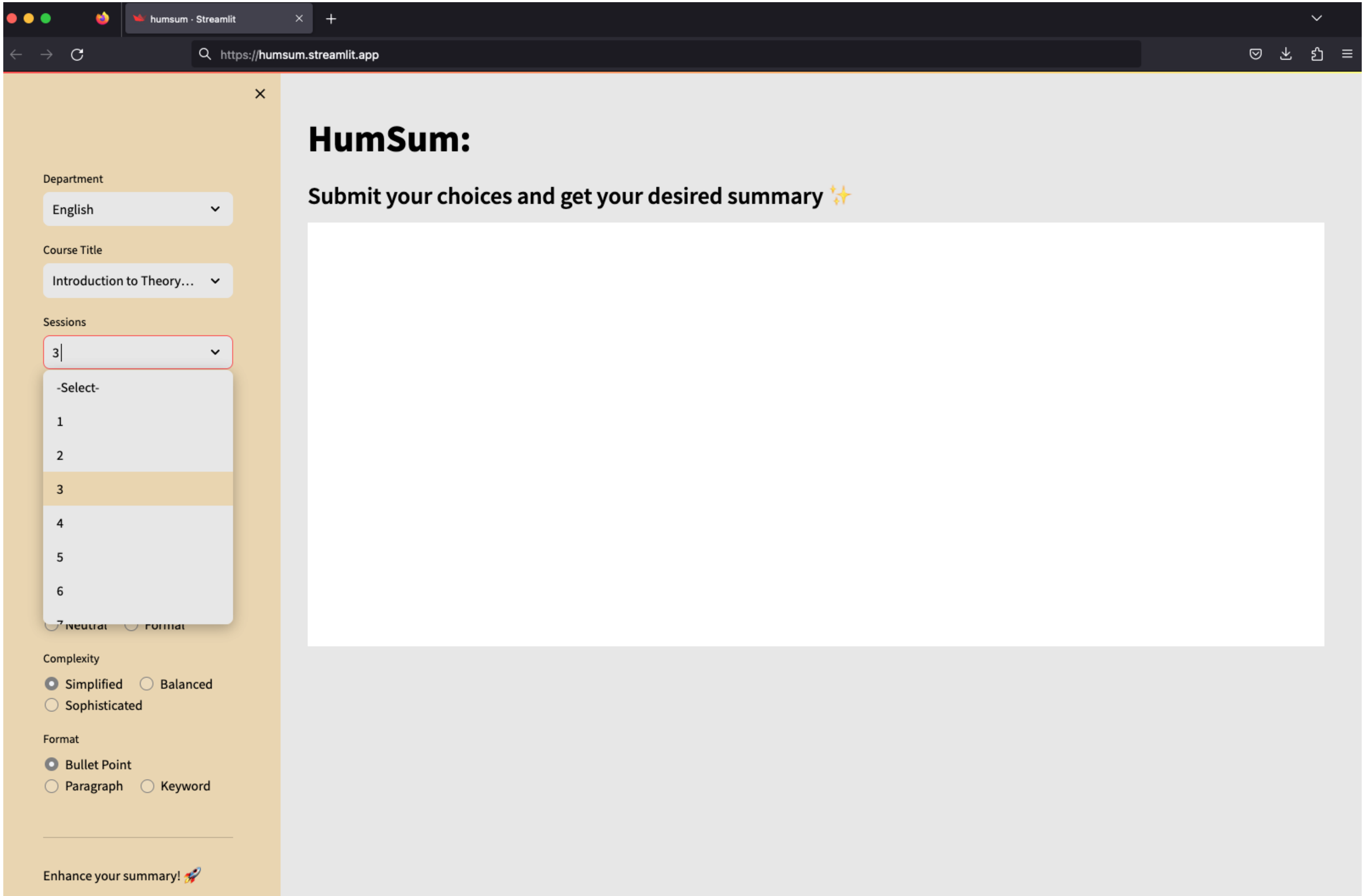

Figure 18: User can select the desired session based on the lecture sessions extracted from OpenYale courses described in section 3.

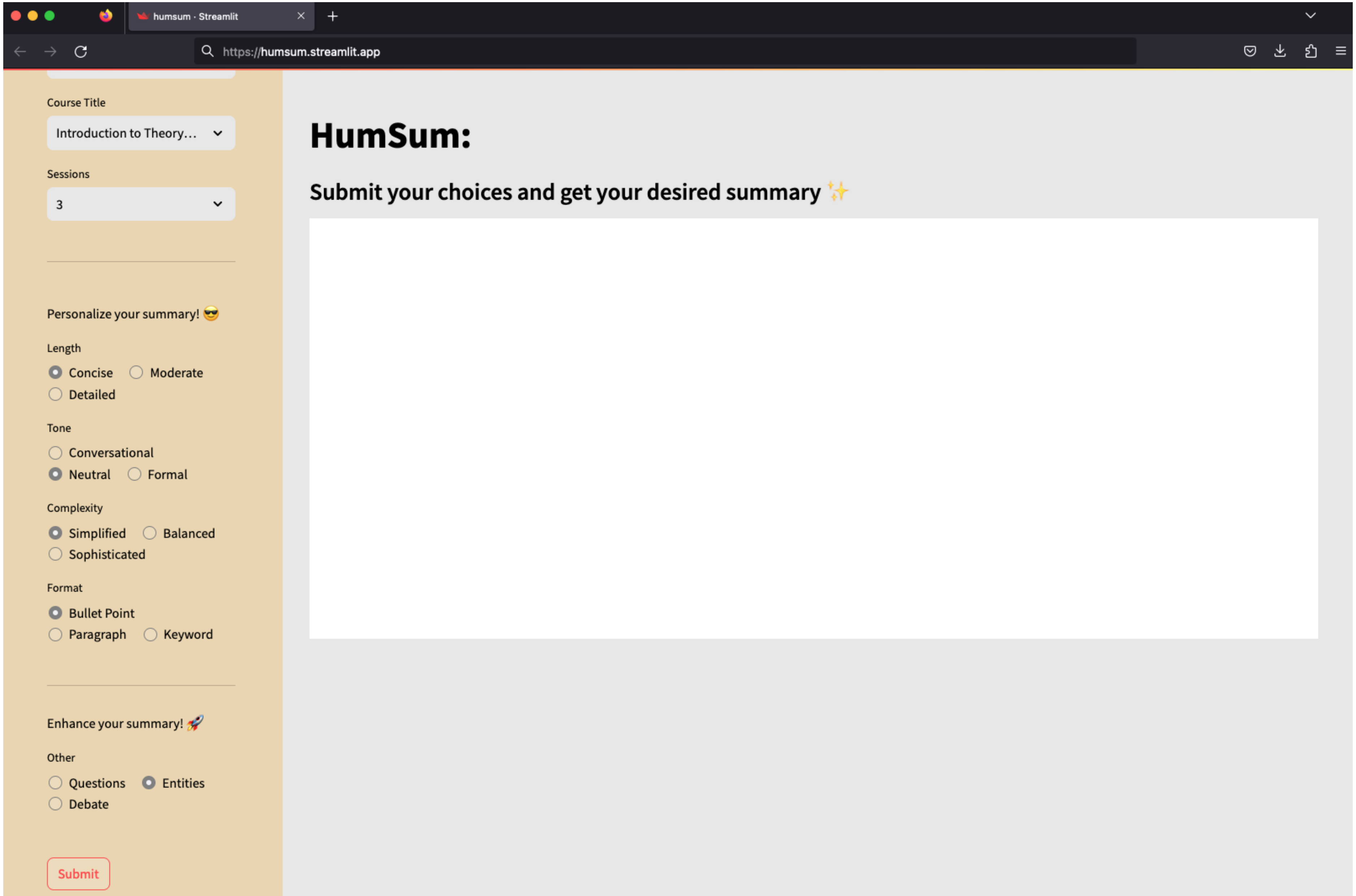

Figure 19: Based on the instruction, the user selected the options as shown in the image and clicked on the submit button to get their personalize summary.

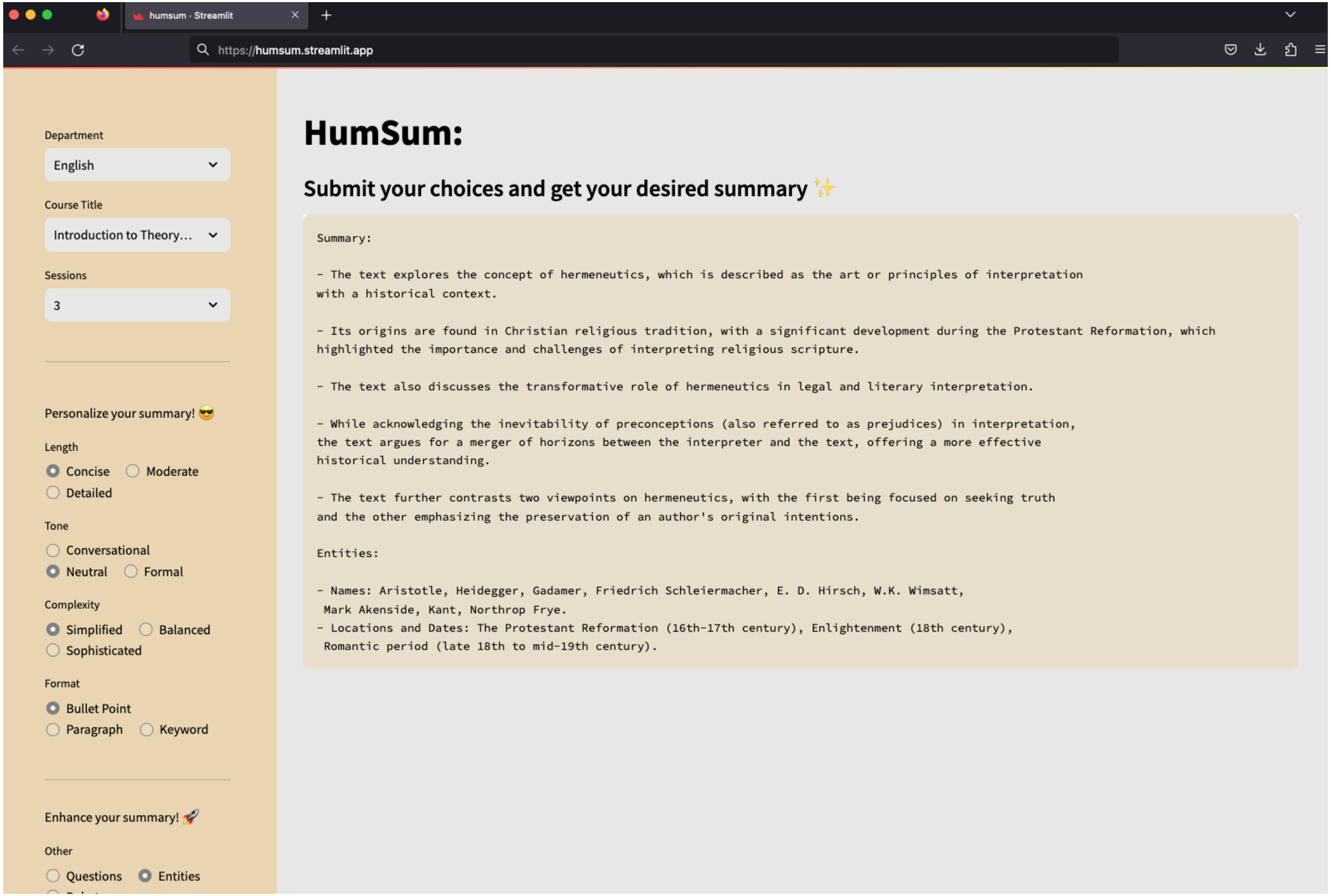

Figure 20: The generated summary based on the instruction using OpenAI's GPT-4-8k model.

A.4 Second Survey Assessment and Evaluation of the Tool

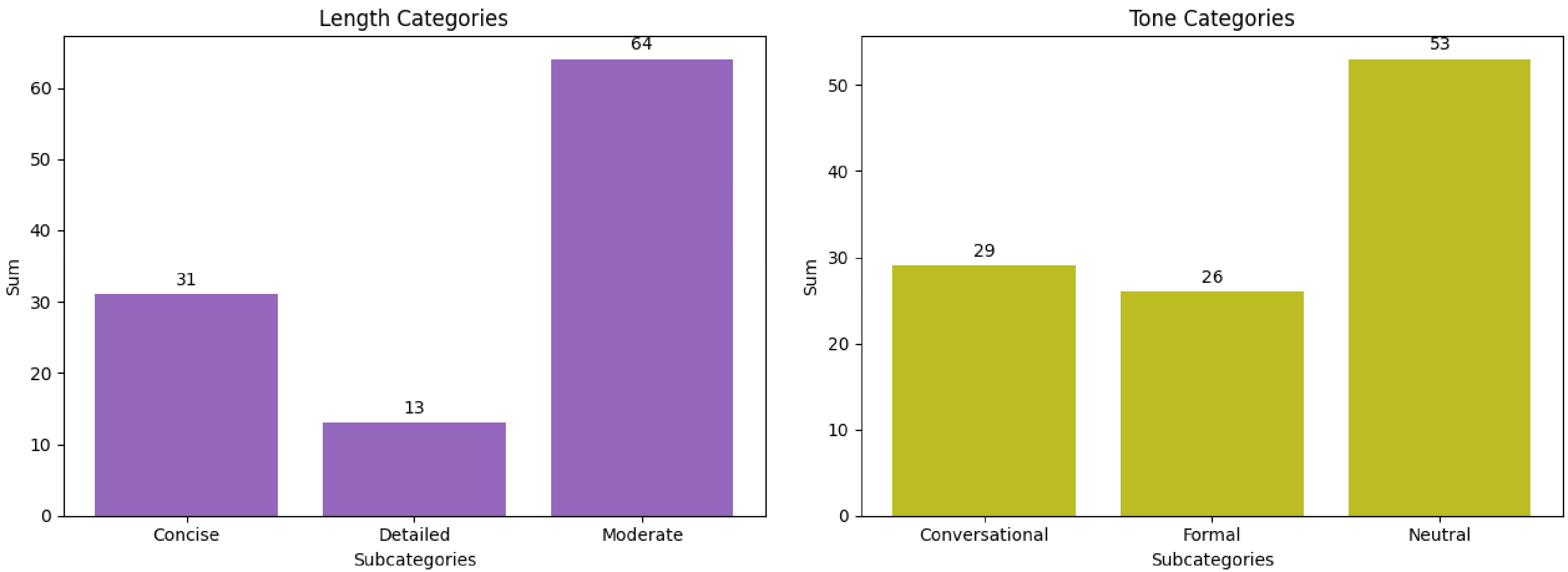

Figure 21: The figures showcase the number of times an item was selected by the user for length and tone options.

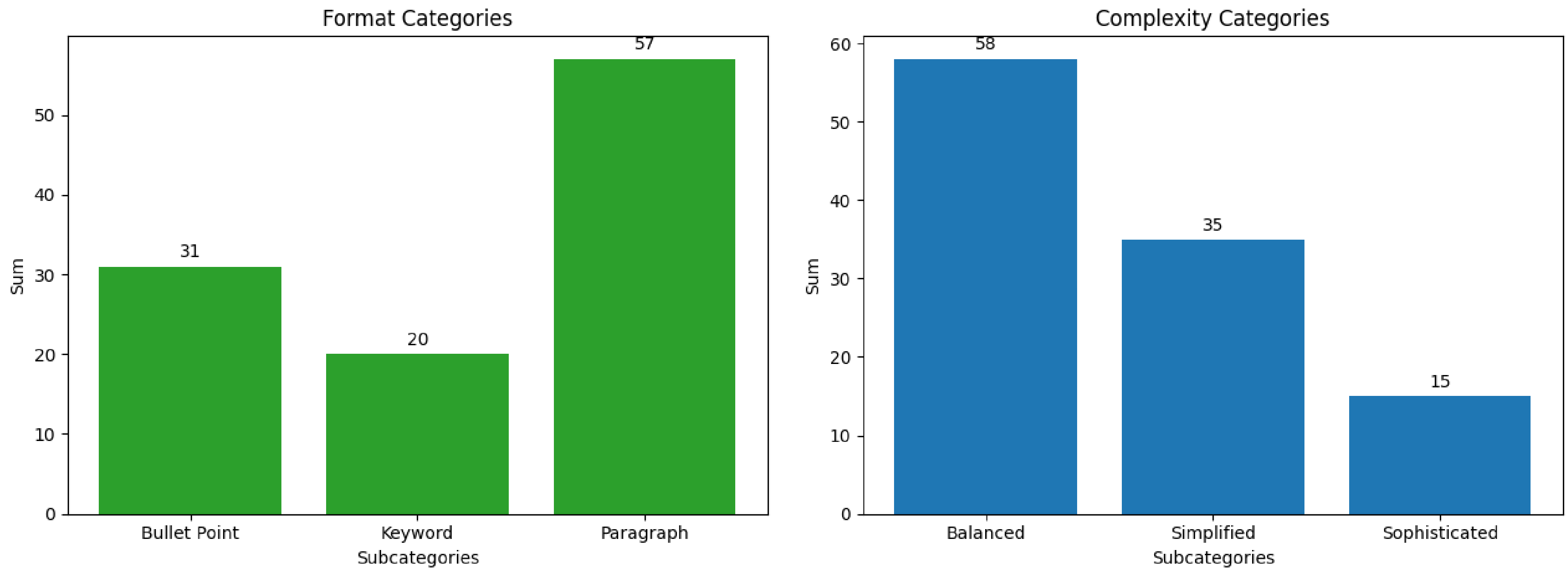

Figure 22: The figures showcase the number of times an item was selected by the user for format and complexity options.

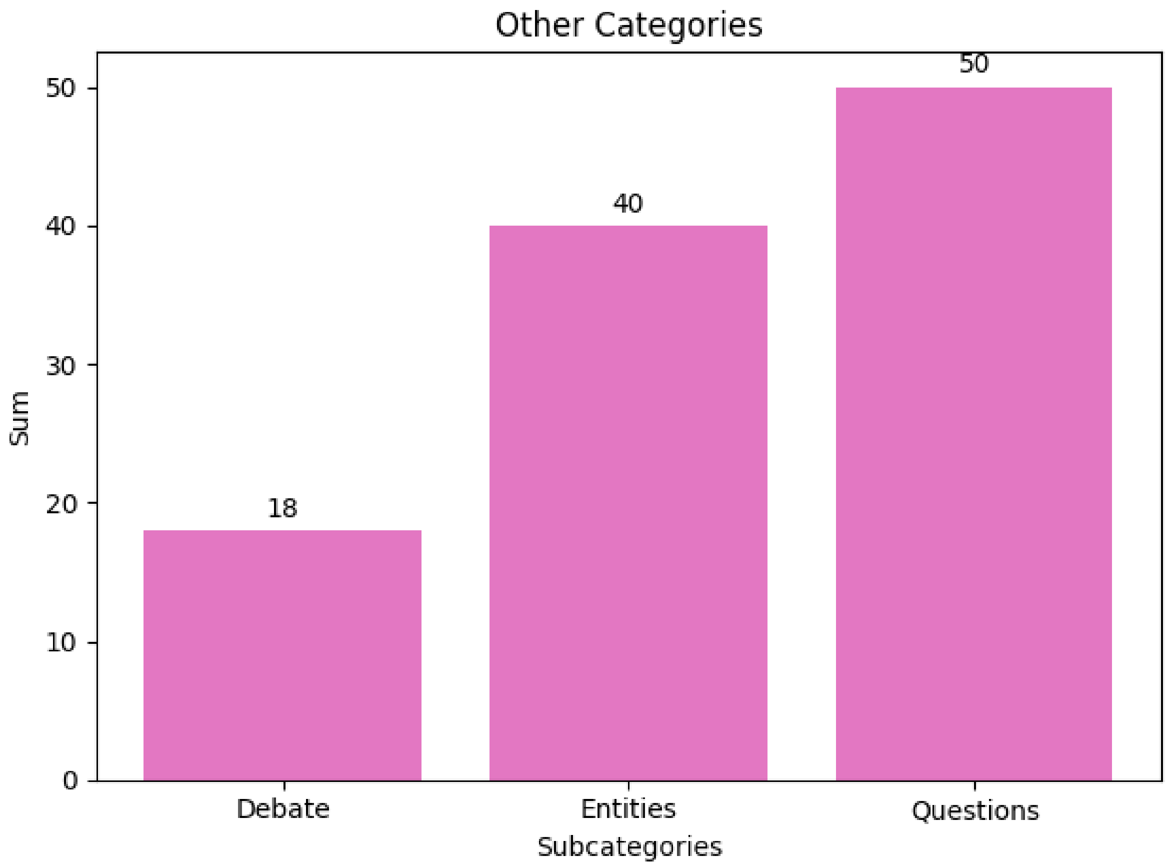

Figure 23: The figures showcase the number of times an item was selected by the user for extra features.

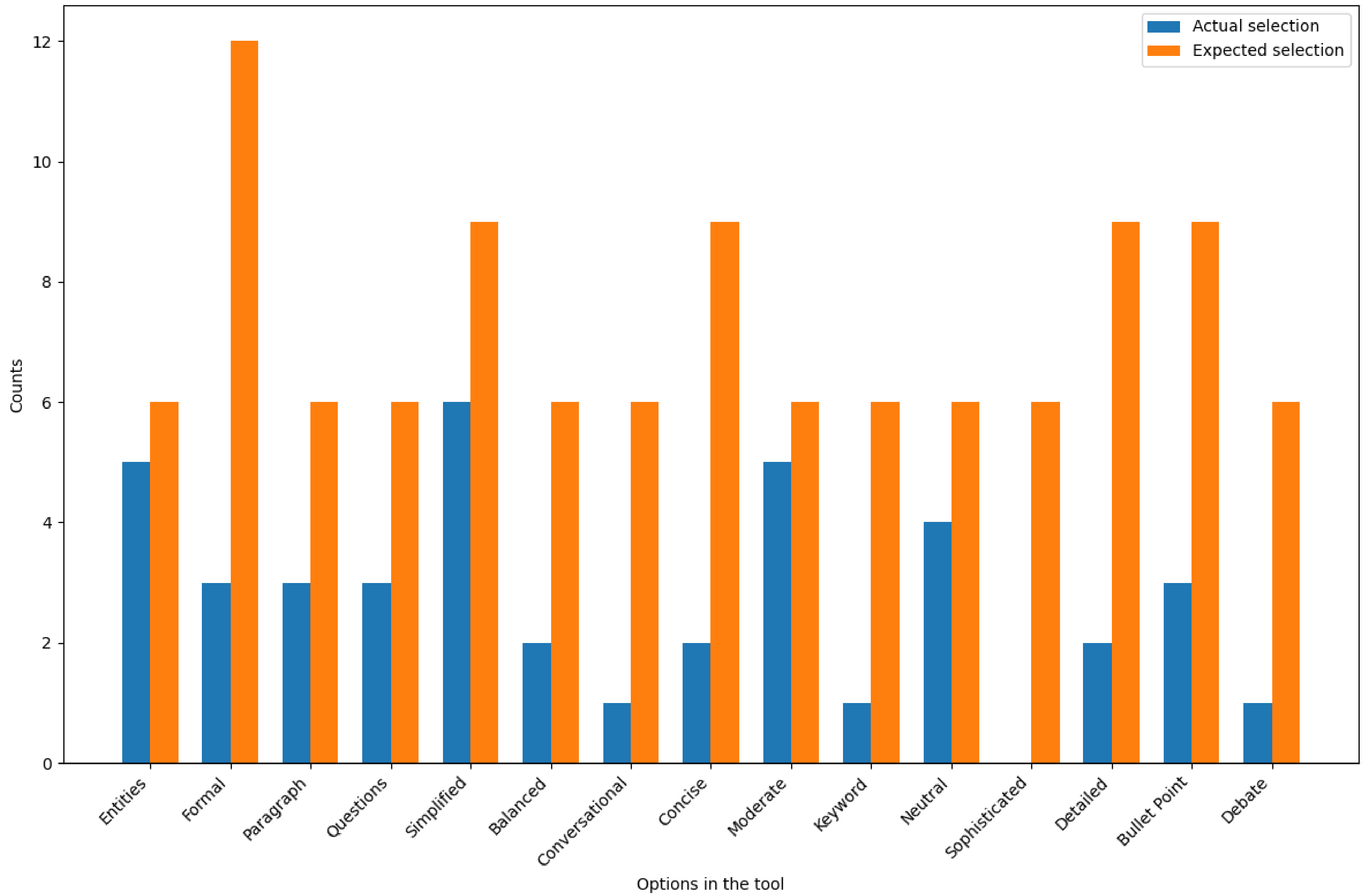

Figure 24: Comparison between our anticipated selections and the actual number of user preferences that aligned with our expectations.

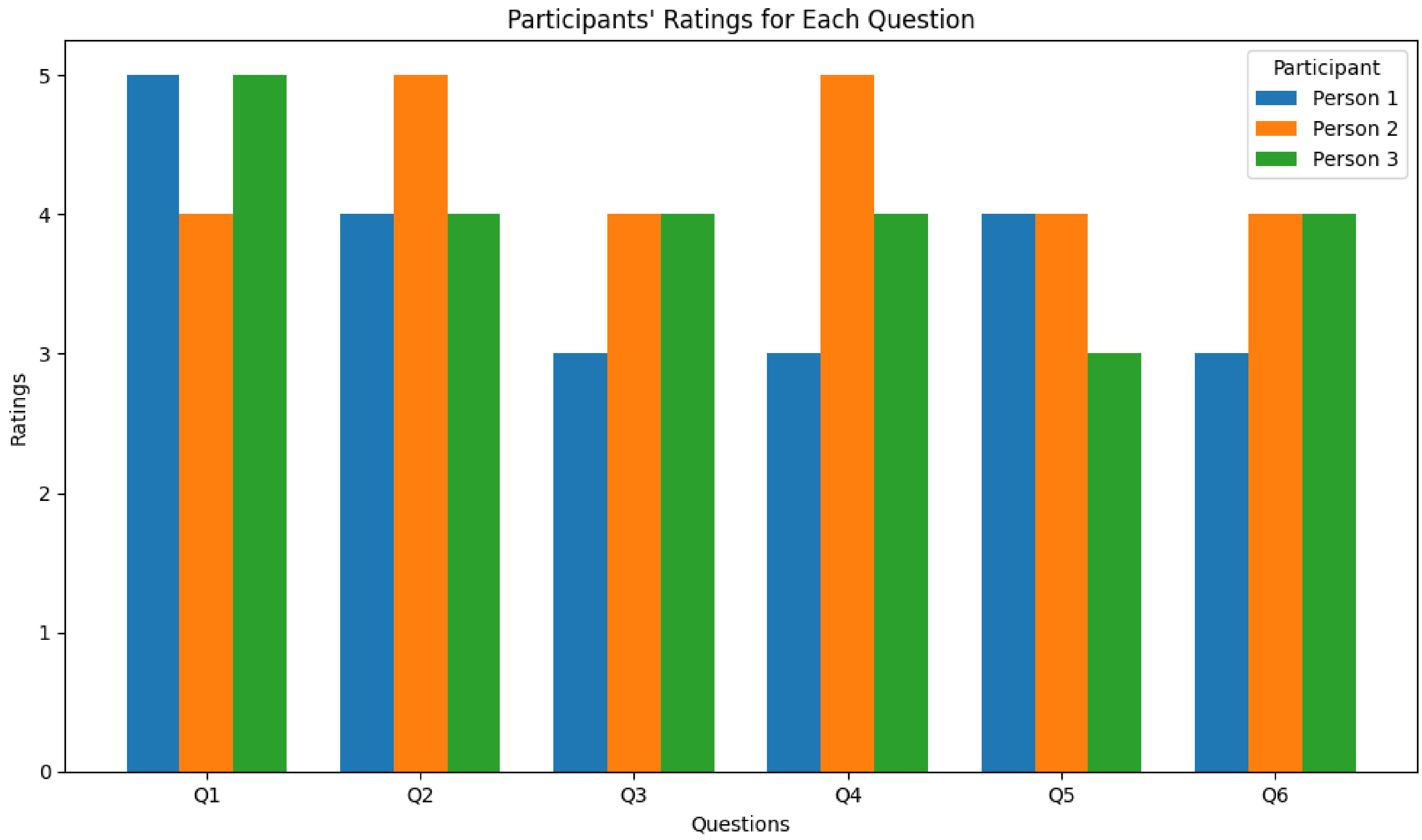

Figure 25: Evaluation of the tool using the final questionnaire encompassing six questions.