Usability testing and technological evaluation results

M18: 30 June 2024

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Grant Agreement nº	101087451
Project acronym	AI4EDU
Project title	Conversational AI assistant for teaching and learning
Funding Scheme	ERASMUS-EDU-2022-PI-FORWARD
Project Duration	01/01/2023 – 31/12/2025 (36 months)
Coordinator	Athina-Erevnitiko Kentro Kainotomias Stis Technologies Tis Pliroforias, Tonepikoinonion Kai Tis Gnosis (ARC)
Associated Beneficiaries	▪ LULEA TEKNISKA UNIVERSITET (LTU) ▪ ELLINOGERMANIKI AGOGI SCHOLI PANAGEA SAVVA AE (EA) ▪ PAIDAGOGIKO INSTITOUTO KYPROU (CPI) ▪ UNIVERSITY OF CYPRUS (UCY) ▪ MANAGEMENT COMMITTEE OF DRUMCONDRA EDUCATION CENTRE (DEC)

Document identifier: D4.2

Version: 1.0

Authors: Gregory Milopoulos (EA), Nikos Pantelaios (EA)

Reviewers: Theodoros Karafylidis (UCY), Spyridoula Stamouli (ARC), Anna Vacalopoulou (ARC)

Dissemination status: PU

Project no. 101087451
AI4EDU

Co-funded by
the European Union

Conversational AI assistant for teaching and learning

ERASMUS-EDU-2022-PI-FORWARD

Start date of project: 01/01/2023

Duration: 36 months

History Chart				
Issue	Date	Changed page(s)	Cause of change	Implemented by
0.1	15.06.2024	-	Final Draft	Gregory Milopoulos
0.2	24.06.2024	All	Changes by reviewers	Reviewers
1.0	26.06.2024	All	Final editing	Gregory Milopoulos

Validation			
No.	Action	Beneficiary	Date
1	Prepared	EA	15.06.2024
2	Approved	ARC	26.06.2024
3	Released	ARC	30.06.2024

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Table of Contents

Co-funded by
the European Union

Co-funded by
the European Union

Executive Summary

This deliverable reports on the results of pilot testing experiments conducted for the first prototypes of the AI4EDU applications: the student's conversational AI assistant – Study Buddy (SB) and the teacher's assistant – Teacher Mate (TM). The evaluation focused on user experience, usability, and technology acceptance, with participation from all partners in the four consortium countries: **Cyprus, Greece**, **Ireland**, and **Sweden**. Following the evaluation methodology defined under Task T4.1 Evaluation of usability and technology acceptance methodology and reported in **D4.1 Evaluation Methodology**, these pilot testing experiments have been successfully completed in all participating countries. This document is structured into six main sections as follows:

- **Section 1 - Introduction**: Provides an overview of AI4EDU, its objectives, and the tools being tested.
- **Section 2 - Survey Description and Demographics**: Details the demographic information of the study participants, including teachers and students from Cyprus, Greece, Ireland, and Sweden, as well as the pilots' implementation in the four project countries.
- **Section 3 – Study Buddy Pilot Results**: Presents the quantitative and qualitative findings from the SB pilot, analyzing user experience, tools functionalities, test completion, chat interactions, and overall usability.
- **Section 4 – Teacher Mate Pilot Results**: Offers a comprehensive analysis of the TM pilot results, covering lesson plan creation, quiz and presentation creation, grading, teaching materials, concept exploration, test generation, and dashboard functionalities.
- **Section 5 - System Data Analysis**: Discusses the system data collected during the pilot studies to further understand user interactions and tool performance.
- **Section 6 – Conclusions & Recommendations**: The conclusions and recommendations are drawn from extensive quantitative and qualitative research conducted across various educational contexts in Cyprus, Greece, Ireland, and Sweden.

Acronyms and abbreviations

Abbreviation	Description
AI4EDU	Artificial Intelligence for Education (project name)
AI	Artificial Intelligence
SB	Study Buddy
TM	Teacher Mate
RAG	Retrieval Augmented Generation

Co-funded by
the European Union

1. Introduction

AI4EDU aims to leverage the power of AI to enhance educational experiences through innovative tools tailored for both teachers and students. This document, **D4.2 Usability Testing & technological evaluation experiments**, presents the findings from usability testing conducted for the AI4EDU applications, **Teacher Mate (TM)** and **Study Buddy (SB)**, in the four project countries. These tests were part of the first cycle of end-user pilots, aimed at evaluating the initial prototype of the AI4EDU tools. Designed to support educators and learners, these tools provide AI-driven assistance in lesson planning, quiz creation, grading, study support, and more. The primary objective of this pilot study was to gather detailed feedback from end users —teachers and students— regarding the usability and technology acceptance of the AI4EDU tools. This feedback will be exploited to make the necessary revisions of the TM and SB applications -as part of the WP5 activities- and develop the final versions of the TM and SB.

SB aims to enhance the learning experience of students by offering personalized and interactive support. This tool's various functionalities are intended to make studying more engaging and effective, thereby improving overall academic performance. SB interacts with students in natural language explaining concepts, answering questions, summarizing texts, extracting key points and providing study tips. SB offers tailored functionalities such us feedback on work as a Mentor, direct instruction as a Tutor, and an efficient assessment experience through its Test Interface. It integrates educational content to ensure curriculum-aligned support. The platform's web-based Test Interface offers a secure and intuitive way to conduct assessments, benefiting both students and educators.

TM brings to educators a set of tools that promise to significantly improve the teaching and learning experience by leveraging prompt engineering, AI, and textbook integration. Central to the TM platform is the Teacher Dashboard, an intuitive interface that enables educators to closely monitor student progress, administer and grade assessments, and deliver focused feedback. The platform uses advanced prompt engineering techniques to assist with lesson planning, material preparation, test creation, and automated grading, thus reducing educators' workloads. Its chat interface, powered by -but not limited to- GPT-3.5 or GPT-4, aids in developing teaching materials and simplifying complex concepts through interactive communication, enhanced with multilingual voice capabilities. A standout feature is its Retrieval Augmented Generation (RAG), which integrates course textbooks into lesson plans, quizzes and other teaching materials, ensuring alignment with the curriculum. TM also offers customizable tools for creating tests and presentations, along with automatic grading to streamline evaluations. The platform supports concept exploration through interactive dialogues and prompts, leveraging best pedagogical practices to cater to diverse learning styles.

Both tools have been tested and evaluated in the context of various use cases of the teaching and learning practice for selected school subjects of Science and Humanities per country, as reported in D2.1 Pedagogical Framework and user requirements. The feedback received from the pilot studies conducted across the project countries is critical for refining and improving the SB and TM tools to better meet the needs of educators and learners. The study also aimed to understand the specific requirements and challenges faced by users in

Co-funded by
the European Union

different educational contexts, ensuring the tools' relevance and effectiveness. The pilot testing followed a structured evaluation methodology defined in Task T4.1 and reported in D4.1, with experiments involving participants -students and teachers- from all partners in the four consortium countries: **Cyprus**, **Greece, Ireland**, and **Sweden**. This comprehensive evaluation and testing approach ensured that the feedback collected was diverse and representative of different educational systems and contexts.

The results from the pilot testing indicated that teachers found TM helpful in streamlining tasks and reducing administrative burdens. However, suggestions for user interface improvements were noted. Students appreciated SB's support in learning, but feedback highlighted the need for clearer explanations and more accurate content generation.

Overall, the pilot testing provided valuable insights into the usability and acceptance of the AI4EDU tools. These findings will guide the refinement of TM and SB, enhancing their potential to support effective teaching and learning practices across diverse educational settings. The successful completion of these pilot studies in all consortium countries demonstrates the tools' adaptability and relevance in various educational contexts, paving the way for broader implementation and impact.

Co-funded by
the European Union

2. Survey Description and Demographics

2.1. Participants

The study participants comprise 32 secondary school teachers and 92 students from **Cyprus, Greece**, **Ireland**, and **Sweden**, ensuring a diverse representation of educational contexts and experiences across these countries (Table 1):

- **Cyprus:** 11 teachers of Biology and History and 10 students, aged 13-15 years.
- **Greece**: 10 teachers of a wide range of subjects, and 57 students, aged 12 and 13 years.
- **Ireland:** 8 teachers of Science and Humanities subjects and 17 students, aged from 13 to 18 years, covering primary to senior secondary education.
- **Sweden:** 3 teachers of Social Sciences, Biology and History and 8 students, aged 17 to 18 years.

This diverse age range of students and extensive teaching experience, with many educators having over two decades in the profession (see 2.1.3), provide a robust foundation for evaluating and enhancing the AI4EDU tools.

Table 1: Pilots participants

Country	N Teachers	N Students
Cyprus	11	10
Greece	10	57
Ireland	8	17
Sweden	3	8
TOTAL	**32**	**92**

2.1.1. Participating Students

The demographic data of the pilot study participants provide a comprehensive overview of the age distribution of students across **Cyprus**, **Greece, Ireland**, and **Sweden**, crucial for understanding their educational contexts and developmental stages.

Cyprus contributes 10 students from the *Aglantzia Gymnasium and from the Forum Private School*, mainly aged **14**, suggesting a focus on mid-secondary education, a transitional phase preparing for higher education and specialized subjects.

In **Greece**, the dataset includes 57 students of the *Ellinogermaniki Agogi private school* in Athens. Students predominantly aged **12** and **13**, indicating early secondary education, characterized by foundational subjects and adjustment to more rigorous demands.

In **Ireland** 16 students from the Junior and Senior Cycle of secondary education, who were attending *Saint Thomas' Senior National School* participated in pilots. Students were aged

Co-funded by
the European Union

12 to 18, with many aged 18, reflecting a broader spectrum from early to late secondary education and highlighting the transition to higher education or vocational paths.

In Sweden, 8 students aged 17 and 18 participated in pilots. The students were attending *Björknäsgymnasiet*, located in the city of Boden. Participating students represent the latter stages of secondary education, dealing with advanced subjects and preparation for tertiary education or the workforce.

Overall, the diverse age range across the four countries, from early to late secondary education, emphasizes the importance of considering age-related factors in tool development to ensure relevance and effectiveness in diverse educational settings.

2.1.2. Participating Teachers

The demographic data of the research encompass a diverse group of teachers from Cyprus, Greece, Ireland, and Sweden, covering a wide range of subjects and educational levels.

In Cyprus the participants were teachers of both lower and higher secondary education in several schools, such as *Aglantzia Gymnasium, Soleas Lyceum and Acropolis Lyceum,* Nicosia, as well as collaborating teachers from the *Cyprus Pedagogical Institute*. They were teaching Biology and History alongside Greek Language. Their teaching spans from lower to upper secondary levels, like their Greek counterparts. These educators generally handle multiple classes, with many teaching more than five classes. Experience among Cypriot teachers varies, with some having 11-20 years of teaching experience, while others have over 21 years. The History and Greek Language teachers demonstrate a significant breadth of experience, which suggests a strong grasp of both subject content and pedagogical skills.

In Greece, the participants were all teachers of the *Ellinogermaniki Agogi* private school in Athens, teaching various subjects of the Greek secondary education curriculum, such as Physics, IT, Mathematics, Biology, History, Greek Language, and English. The class grades they teach range from lower secondary (7th grade) to upper secondary (12th grade), with most teachers managing more than five classes simultaneously. Most of these educators have extensive teaching experience, with several having taught for over 20 years. Specifically, the Physics and IT teachers have been in the profession for 21-30 years, reflecting a deep reservoir of experience and expertise. However, even those with less experience, such as those teaching History and Greek Language, have between 6-10 years of experience, indicating a solid foundation in educational practice.

In Ireland, the participating teachers were teaching at five different secondary schools in Dublin, namely *Lucan East Educate Together NS, St. Setons Secondary School, St. Marks. SNS, St. Thomas SNS, Griffeen and Kishogue Community Colleges*. Irish teachers contribute to the diversity of this research by teaching a broad spectrum of subjects, including Environmental Studies, Sciences, Geography, Science, Mathematics, and Technology. They teach across all grade levels, from primary education to both junior and senior secondary cycles. Irish educators, too, tend to manage more than five classes, with a considerable number boasting over 30 years of teaching experience. This extensive experience is especially pronounced among Science and Environmental Studies teachers, indicating a deep-seated commitment to education and a wealth of knowledge in these fields.

Co-funded by
the European Union

Lastly, teachers from the *Björknäsgymnasiet* located in the city of Boden, **Sweden,** focus on subjects such as Entrepreneurship, Building Automation, English, Swedish, and Mathematics, within the secondary school level. They generally handle between three to five classes. The teaching experience among Swedish educators is substantial, with many having 21-30 years of experience.

In summary, the demographic data of the participating teachers showcases a highly experienced and diverse group of educators across all four countries. These teachers cover a wide range of subjects and educational levels, demonstrating a broad spectrum of teaching expertise and pedagogical skills. The majority manage multiple classes and possess extensive teaching experience, with many having over two decades in the profession. This diversity and depth of experience among the participating teachers provide a robust foundation for understanding, evaluating and enhancing TM, ensuring it meets the needs of educators across various contexts and educational systems.

2.2. Study Buddy and Teacher Mate pilots' implementation

The Study Buddy and Teacher Mate pilots, aiming to gather user feedback on the usability and technology acceptance of the AI4EDU tools in educational settings, were organized and implemented by the project partners' teams across the four project countries.

The pilot workshops in **Cyprus** were organised and implemented by the project partners University of Cyprus and Cyprus Pedagogical Institute and were conducted at the Cyprus Pedagogical Institute in Nicosia. Both student and teacher pilots took place on *April 20, 2024*.

In **Greece**, the pilots were hosted by the Ellinogermaniki Agogi private school in Athens, the project partner that also organised and implemented student and teacher pilots. The student pilots occurred between *April 11 and 22, 2024*, while the teacher pilots were conducted between *May 24 and June 5, 2024*.

In **Ireland**, the partner who organised and implemented the pilots was Drumcondra Education Center, Dublin. The student pilot was held on *March 15, 2024*, at Saint Thomas' SNS in Dublin. The teacher pilot took place on *March 14, 2024*, at the Griffeen Community College and Kishoge Community College, both located in Dublin.

In **Sweden**, Luleå University of Technology organized and implemented the pilot workshops. The student pilot was conducted on April 10, 2024, and the teacher pilot on *March 27, 2024*, both at the Björknäsgymnasiet in the city of Boden.

During pilots implementation, the project team of Athena Research Center provided continuous technical support to ensure that any issues encountered by users were promptly addressed, enabling a smooth and productive user experience.

Co-funded by
the European Union

Bellow, there are some indicative photos from students and teachers pilot workshops, as conducted in the four project countries:






2.3. Methodology

The methodology section outlines the structured protocol for the TM and SB pilots, providing clear steps for the team members conducting the pilots, as described in D4.1. The first step was to obtain **informed consent** from teachers, as well as from parents or guardians and students. Participants were given time to review these documents and ask questions before signing them, ensuring informed participation.

The pilot workshops for testing and evaluating TM and SB followed a comprehensive methodology involving demonstrations of each tool and functionality of the platform, hands-on practice, and detailed feedback collection through online questionnaires, which contained both closed-ended and open-ended questions, to ensure comprehensive quantitative and qualitative data collection on usability and technology acceptance. The questionnaires addressed to students and teachers were available in English for Ireland and Sweden, and in Greek for Cyprus and Greece.

The implementation of the pilot workshops followed the same methodology across countries. Team members guided participants through platform login and account creation, demonstrated various functionalities of SB and TM and facilitated the use of the tools by the end users, responding to questions or providing help on the spot.

2.3.1. Study Buddy Pilot Methodology

Initially, participating students and their parents or guardians received a Participant **Information Sheet and Informed Consent Form**, which they read and signed, while the project team members addressed any questions they had.

Participants then created student accounts with specific username formats to facilitate data analysis. They accessed the platform at https://apps.ilsp.gr/ai4edu/, they read and agreed to the Terms of Use and Privacy Statement.

Next, participants filled in Section 1: "Participant's Information" of the questionnaire using the provided links to the questionnaire of the respective country.

The protocol for testing and evaluating the SB platform was structured as part of the D4.1 "Methodology for the evaluation of usability and technology acceptance". However, some modifications were made to follow the latest features added to the SB and TM applications (see Appendix). The protocol included three main steps:

A. **Demonstration:** Each SB tool was presented by project team members, emphasizing its usefulness in specific use-case scenarios. Parameters for prompt creation were explained per tool, and enough time was allocated to review and comment on SB's responses, displayed on a shared presentation screen. The responses were then discussed with students to ensure understanding of the tools' functionalities and

Co-funded by
the European Union

their practical applications. The demonstration covered both the general chat and textbook-specific chat functionalities.

B. **Hands on practice:** Students actively engaged with the demonstrated tools under the guidance of project team members/facilitators. The project team members ensured students had sufficient time to familiarize themselves with the tools, offered assistance for any practical challenges encountered, and provided suggestions for creating prompts. Emphasis was placed on practical usage scenarios such as self-study, homework, and test preparation in selected Science and Humanities subjects for each country. Participants were instructed to utilize both the general chat and textbook-specific chat options during the session.

C. **Feedback collection** through an online questionnaire for students (see Appendix): As soon as the demonstration and the hands-on practice sessions were completed, participants evaluated the respective tools and functionalities of SB, one by one. Participants were given ample time to fill in all questions within each section of the questionnaire. Moreover, they were encouraged to respond to open questions to provide qualitative feedback on the actual use of the tools.

Following the steps outlined above, the project team members/facilitators demonstrated and guided students in using the following tools:

i. **Explain Term Tool**: Demonstrated by typing a term, showing both the General and Textbook-specific Chat functionalities. Students then independently used the tool, noting any mistakes.
ii. **Interactive Concept Learning**: Facilitators demonstrated its usage, followed by students practicing with their own prompts.
iii. **Summarize Text and Extract Key Points**: Text was copied from educational sources and pasted into the tool for demonstration purposes. Then students practiced using their own texts as input for summarization.
iv. **Assessment Quiz**: Demonstrated for test preparation purposes, after which students created their own quizzes to test their understanding.
v. **Revise Written Assignment**: Drafts of written assignments were pasted into the tool. Parameters for prompt creation were explained, and students practiced revising their own assignments using the tool.
vi. **Grading**: Demonstrated to help students understand their mistakes and improve assignments, then students practiced.

Subsequently, students filled in **Section 2 of the questionnaire**, to evaluate all the above tools' usability and technology acceptance collectively.

Additionally, the following SB tools and functionalities were demonstrated and evaluated during the pilot:

vii. **Test Tool and Dashboard:** Teachers first generated and assigned a test to students using TM. Students completed the test, and then both students and teachers

Co-funded by
the European Union

reviewed the results in their respective dashboards. Participants evaluated both the test and the dashboard's functionalities.

viii. **Chat Functionality and Chat History:** The demonstration included interactions with SB, encompassing both text-based and voice-based interactions. Participants actively engaged in these interactions and explored the chat history to assess the quality of responses provided by the SB.

After each of the above tools' demonstrations and hands-on practice sessions, students were encouraged to evaluate them, provide feedback on any issues encountered and suggestions for improvement, by completing **Sections 3, 4 and 5 of the questionnaire**. Finally, they filled in **Section 6** of the questionnaire, containing general usability and technology acceptance questions.

Observations of the testing and evaluation workshops were gathered by the project teams who conducted the pilots and were meticulously documented in an **Observation sheet** to guide improvements in the tools.

2.3.2. Teacher Mate Pilot Methodology

The protocol for testing and evaluating TM followed the same steps of the SB pilot protocol (see section 2.3.1.) and involved a structured demonstration of each tool within TM, followed by hands-on practice of each demonstrated tool and feedback collection regarding the evaluation of each tool through an online questionnaire designed for teachers (see Appendix for both protocol and questionnaire).

Participants first read and signed the informed consent form and created their own accounts to log into the platform, following a specific username format for easy identification of the log files. They then filled out a preliminary questionnaire section regarding their demographic information.

The teachers tested and evaluated the following TM tools and functionalities:

i. **Lesson Plan Creation:** Teachers were shown how to use the lesson plan tool using both general chat and textbook-specific prompts. Teachers created two lesson plans, one through general chat and another via textbook-specific prompts. Participants then completed related questionnaire sections.

ii. **Quiz Creation, Presentation Creation, Grading, Teaching Materials, and Concept Exploration:** Each tool was demonstrated individually, and then used by the participants. Emphasis was placed on using both general chat and textbook-specific options. After using each tool, participants filled out corresponding questionnaire sections.

iii. **Test Generation and Class Management:** Teachers were guided through the parameters and process of test generation. Teachers then created a class, generated a test, and assigned it to students. Participants acted as students, completed the test, and then reviewed results in the Teacher Dashboard.

Co-funded by
the European Union

iv. **Chat Functionality and Chat History:** The functionality of TM's chat, including text and voice interactions, was demonstrated. Participants engaged in chat interactions and explored the chat history to assess response quality.

Throughout the pilot, participants were consistently encouraged to be attentive to any issues encountered, and complete detailed questionnaire sections after each activity to provide quantitative and qualitative feedback. Observations and feedback were gathered by the project teams who conducted the pilots and were documented in an **Observation sheet** to guide improvements in the tools.

3. Study Buddy Pilot results

3.1. Quantitative Analysis

In the following subsections, the quantitative data collected by students through questionnaires are analysed. The subsections follow the student questionnaire structure:

- **Questionnaire section 2**: Experience using the Study Buddy tools (Explain a term, Interactive concept learning, Summarize text, Extract key points, Assessment quiz, Revise written assignment, Grading)
- **Questionnaire section 3**: Test taking and Student Dashboard
- **Questionnaire section 4**: Chat with Study Buddy
- **Questionnaire section 5**: Chat history and evaluation of quality of SB responses
- **Questionnaire section 6**: General usability and technology acceptance questions

Each section contained closed-ended questions aiming to collect information of the SB tool's effectiveness, usability, and areas for enhancement, ensuring they meet the diverse needs of students.

At the end of each section of the questionnaire, students were asked to share any additional observations or recommendations about the SB tools and functionalities. This open-ended feedback provided a platform for users to express their thoughts, experiences, and suggestions on revisions, new features or new functionalities to be added, offering valuable insights for continuous improvement. The feedback provided by students in these open-ended questions will be presented in **Section 3.2, Qualitative Analysis**, which supplements the quantitative analysis presented here.

3.1.1. Experience Using the Study Buddy Tools

This section is dedicated to evaluating the user experience, the usability and the acceptance of the SB tools, namely **Explain Term, Interactive concept learning, Summarize Text, Extract key points, Assessment quiz, Revise written assignment,** and **Grading**. It aims to gather detailed insights on various aspects of these educational tools to understand their impact and identify areas for improvement.

Co-funded by
the European Union

The questions cover a range of topics, starting with the **usefulness** of the explanations provided by the SB tools, including examples, summaries, key points, quizzes, and feedback for daily homework (4a). This question assesses how well the tools can support students in their daily academic tasks.

Next, the **ease** of using the various SB tools is examined (4b), focusing on the user-friendliness and accessibility. Understanding how easily students can navigate and utilize these features is crucial for ensuring a smooth user experience.

The extent to which students think that the SB Tools **enhance their knowledge** of the subjects they are learning is also explored (4c). This question aims to determine the tools' effectiveness in supporting and enriching students' learning processes.

Furthermore, the **perceived impact** of SB tools on the quality and effectiveness of everyday study is assessed (4d). This question evaluates whether students think that the tools can help them study more efficiently and effectively.

Additionally, the tools' role in **empowering** students become more efficient users of Generative AI is considered (4e).

The section also examines the extent to which students think the SB tools can contribute to improving their **performance at school** (4f). This question seeks to understand the broader academic benefits of using the above tools.

3.1.1.1. Usefulness of the Study Buddy Explanations (Question 4a)

The usefulness of the Study Buddy (SB) explanations presented some variations across the surveyed countries:

- **Cyprus**: Students unanimously found the SB tools useful, with 70% finding them very useful and 30% finding them extremely useful. This indicates a strong acceptance and perceived benefit of the tools in Cyprus.
- **Greece**: Feedback was mixed, with 35.09% finding the tools very useful, 10.53% extremely useful, but also 31.58% moderately useful and 22.81% expressing a negative view. This suggests that while some students found the tools helpful, others were either unsure or dissatisfied.
- **Ireland**: Responses were very positive, with 47.06% finding the tools very useful and 35.29% extremely useful. This high level of satisfaction suggests that the tools are well-received and considered valuable by Irish students.
- **Sweden**: Most students (62.5%) found the tools very useful, though 25% found them slightly useful. This shows a generally positive reception, but a significant minority found the tools lacking.

In summary, while Cyprus and Ireland showed high satisfaction, Greece presented a more mixed response, indicating areas for improvement. Overall, the majority of students found the SB tools useful.

Co-funded by
the European Union

Table 2: 4a. How useful did you find the explanations of Study Buddy, the examples it provided, the summary and key points it generated, the quizzes it created and the feedback it gave to you for your daily homework?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	30	70	0	0	0	**100,00**	**0,00**	**0,00**
Greece	10,5	35,09	31,58	10,53	12,2	**45,61**	**31,58**	**22,81**
Ireland	35,2	47,06	5,88	11,76	0,00	**82,35**	**5,88**	**11,76**
Sweden	0,00	62,50	12,50	25,00	0,00	**62,50**	**12,50**	**25,00**
Overall	16,3	43,48	21,74	10,87	7,61	**59,78**	**21,74**	**18,48**

3.1.1.2. Ease of Use (Question 4b)

The ease of use of SB tools also demonstrated varied feedback:

- **Cyprus**: 90% of students found the tools easy to use, with 50% finding them very easy and 40% finding them extremely easy. This indicates a high level of user-friendliness.
- **Greece**: 73.68% had a positive view, with 45.61% finding the tools very easy and 28.07% finding them extremely easy. However, 15.79% were moderately easy and 10.53% had negative views, suggesting some difficulties.
- **Ireland**: A high level of satisfaction was reported, with 76.47% positive views, but 17.65% found it difficult to use. This indicates that while the majority found the tools easy to use, a notable minority experienced difficulties.
- **Sweden**: Students unanimously found the tools easy to use (100%), with 75% finding them very easy and 25% finding them extremely easy. This indicates excellent usability in Sweden.

In summary, Cyprus and Sweden reported excellent usability, whereas Greece and Ireland had a notable segment of undecided or negative responses. Overall, the ease of use was well-received by most students.

Table 3: 4b. How easy was it to use the Study Buddy tools?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	40	50	10	0	0	**90,00**	**10,00**	**0,00**
Greece	28,07	45,61	15,79	3,51	7,02	**73,68**	**15,79**	**10,53**
Ireland	52,94	23,53	5,88	17,65	0,00	**76,47**	**5,88**	**17,65**
Sweden	25,00	75,00	0,00	0,00	0,00	**100,00**	**0,00**	**0,00**
Overall	33,70	44,57	11,96	5,43	4,35	**78,26**	**11,96**	**9,78**

Co-funded by the European Union

3.1.1.3. *Educational Impact (Question 4c)*

Students' perceptions of the educational impact of SB tools is reflected in the following results:

- **Cyprus**: 90% had a positive view, with 70% finding the tools very impactful and 20% extremely impactful. This indicates a strong belief in the educational value of the tools.
- **Greece**: Mixed feedback with 42.11% finding the tools very impactful, 7.02% extremely impactful, and significant neutral and negative responses (31.58% and 19.30% respectively). This suggests that while some students saw educational benefits, others did not.
- **Ireland**: 82.35% had positive views, with 47.06% finding the tools extremely impactful. This shows strong support for the educational impact of the tools.
- **Sweden**: 50% had positive views, with 25% finding the tools very impactful and 25% extremely impactful, while the other 50% were neutral. This indicates room for improvement.

In summary, Cyprus and Ireland viewed the educational impact very positively, while Greece and Sweden presented more varied responses. Overall, the educational impact was seen positively by most students.

Table 4: 4c. To what extent do you think that the Study Buddy Tool enables you to know more about the subjects that you are learning?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	20	70	10	0	0	**90,00**	**10,00**	**0,00**
Greece	7,02	42,11	31,58	12,28	7,02	**49,12**	**31,58**	**19,30**
Ireland	47,06	35,29	0,00	11,76	5,88	**82,35**	**0,00**	**17,65**
Sweden	25,00	25,00	50,00	0,00	0,00	**50,00**	**50,00**	**0,00**
Overall	17,39	42,39	25,00	9,78	5,43	**59,78**	**25,00**	**15,22**

3.1.1.4. *Quality and Effectiveness of Study (Question 4d)*

Students' perceptions of the quality and effectiveness of the SB tools is reflected in the following results:

- **Cyprus**: 70% had a positive view, with 30% finding the tools extremely effective and 40% very effective. This indicates a strong belief in the quality and effectiveness of the tools.
- **Greece**: Mixed feedback with 36.84% finding the tools moderately effective, 31.58% very effective, and significant neutral and negative responses (36.84% and 22.81%

Co-funded by the European Union

respectively). This suggests that while some students saw the tools as effective, others did not.

- **Ireland**: 76.47% had positive views, with 41.18% finding the tools extremely effective. This shows strong support for the effectiveness of the tools.
- **Sweden**: 50% had positive views, with 37.50% finding the tools very effective and 12.50% extremely effective, while the other 50% were neutral. This indicates room for improvement.

In summary, Cyprus and Ireland viewed the effectiveness of the tools very positively, while Greece and Sweden presented more varied responses. Overall, the quality and effectiveness were seen positively by most students.

Table 5: 4d. To what extent do you think that the Study Buddy Tools improve the quality and effectiveness of your everyday study?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	40	30	30	0	0	**70,00**	**30,00**	**0,00**
Greece	8,77	31,58	36,84	14,04	8,77	**40,35**	**36,84**	**22,81**
Ireland	41,18	35,29	11,76	11,76	0,00	**76,47**	**11,76**	**11,76**
Sweden	12,50	37,50	37,50	12,50	0,00	**50,00**	**37,50**	**12,50**
Overall	18,48	32,61	31,52	11,96	5,43	**51,09**	**31,52**	**17,39**

3.1.1.5. Efficiency in Using Generative AI (Question 4e)

Students' perceptions of the ability of the SB tool to empower them in using Generative AI are presented in the following results:

- **Cyprus**: 100% had a positive view, with 60% finding the tools very helpful and 40% extremely helpful. This indicates a strong belief that the tools help students become more efficient users of Generative AI.
- **Greece**: Mixed feedback with 36.84% finding the tools moderately helpful, 33.33% very helpful, and significant neutral and negative responses (36.84% and 17.54% respectively). This suggests that while some students saw improvements in their efficiency using Generative AI, others did not.
- **Ireland**: 82.35% had positive views, with 64.71% finding the tools extremely helpful. This shows strong support for the tools' effectiveness in making students more efficient users of Generative AI.
- **Sweden**: 62.50% had positive views, with 37.50% finding the tools very helpful and 25.00% extremely helpful, while the other 37.50% were neutral. This indicates room for improvement.

In summary, Cyprus and Ireland evaluated the tools regarding their potential to help them become more efficient users of Generative AI very positively, while Greece and Sweden presented more varied responses. Overall, the tools were seen positively by most students in enhancing their efficiency with Generative AI.

Co-funded by the European Union

Table 6: 4e. To what extent do you think that using the Study Buddy Tools help you become a more efficient user of Generative Artificial Intelligence?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	40	60	0	0	0	**100,00**	**0,00**	**0,00**
Greece	12,28	33,33	36,84	7,02	10,53	**45,61**	**36,84**	**17,54**
Ireland	64,71	17,65	0,00	11,76	5,88	**82,35**	**0,00**	**17,65**
Sweden	25,00	37,50	37,50	0,00	0,00	**62,50**	**37,50**	**0,00**
Overall	26,09	33,70	26,09	6,52	7,61	**59,78**	**26,09**	**14,13**

3.1.1.6. Performance Improvement at School (Question 4f)

Regarding students' perceptions of the SB tools' impact on their school performance, the results were the following:

- **Cyprus**: 70% had a positive view, with 40% finding the tools very helpful and 30% extremely helpful. This indicates a strong belief in the tools' impact on performance.
- **Greece**: Mixed feedback with 38.60% finding the tools moderately helpful, 31.58% very helpful, and significant neutral and negative responses (38.60% and 21.05% respectively). This suggests that while some students saw an improvement in performance, others did not.
- **Ireland**: 64.71% had positive views, with 47.06% finding the tools extremely helpful. This shows strong support for the tools' impact on performance.
- **Sweden**: 50% had positive views, with 25% finding the tools very helpful and 25% extremely helpful, while the other 50% were neutral or negative. This indicates room for improvement.

In summary, Cyprus and Ireland viewed the tools' impact on performance very positively, while Greece and Sweden presented more varied responses. Overall, the tools were seen as positively impacting performance by most students.

Table 7: 4f. To what extent do you think that using the Study Buddy tools will help you improve your performance at school?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	30	40	30	0	0	**70,00**	**30,00**	**0,00**
Greece	8,77	31,58	38,60	7,02	14,04	**40,35**	**38,60**	**21,05**
Ireland	47,06	17,65	23,53	11,76	0,00	**64,71**	**23,53**	**11,76**
Sweden	25,00	25,00	25,00	12,50	12,50	**50,00**	**25,00**	**25,00**

Co-funded by the European Union

Overall	19,57	29,35	33,70	7,61	9,78	**48,91**	**33,70**	**17,39**

3.1.1.7. Conclusions on the Experience Using the Study Buddy Tools

The feedback on the Study Buddy (SB) tools varied across different countries, highlighting both strengths and areas for improvement.

In **Cyprus** Students unanimously found the SB tools useful, easy to use, and impactful in their educational experience. The tools were particularly praised for helping students become more efficient users of Generative AI. Overall, the tools were highly regarded across all dimensions, with no significant areas of improvement identified.

In **Greece**, the feedback was mixed. While a notable portion of students found the tools useful and easy to use, there were also significant numbers of undecided and negative responses. The educational impact and efficiency in using Generative AI were areas where students expressed varying levels of satisfaction, indicating a need for enhancements in these areas.

In **Ireland** responses were very positive overall, with high satisfaction regarding the usefulness, ease of use, and educational impact of the tools. However, some students experienced difficulties with usability and efficiency in using Generative AI, suggesting these as potential areas for improvement.

In **Sweden**, most students found the tools useful and easy to use, but there was a significant minority who were neutral or negative about the tools' usefulness, educational impact, and efficiency in using Generative AI. These findings suggest that while the tools are generally well-received, there is room for improvement to better meet the needs of all students.

Summary Table: Areas of Improvement and Action points

Table 8: Experience Using the Study Buddy Tools - Summary Table

Country	Areas of Improvement	Action points
Cyprus	None identified	N/A
Greece	Usability, educational impact, efficiency in using Generative AI	Simplify user interface, enhance educational content, provide better training on Generative AI
Ireland	Usability, efficiency in using Generative AI	Improve navigation, offer advanced tutorials on Generative AI
Sweden	Usefulness, educational impact, efficiency in using Generative AI	Tailor content to student needs, improve feedback mechanisms, provide comprehensive AI training

As on **overall conclusion**, The SB tools were well-received in Cyprus and Ireland, with high levels of satisfaction reported. Greece and Sweden had more mixed feedback, highlighting specific areas where the tools could be improved to enhance user experience and educational outcomes.

Co-funded by
the European Union

3.1.2. Test taking and Student Dashboard

This section focuses on the SB Test taking and Dashboard features, aiming to evaluate their effectiveness and user experience. The questions are designed to gather detailed feedback on several aspects related to performing tests and understanding performance metrics. First, the ease of performing teacher-assigned tests using the SB Test Tool is assessed (7a), providing insights into the user-friendliness and accessibility of the testing process.

Following this, the usefulness of the SB Test Tool for completing these assigned tests is examined (7b), highlighting how well the tool supports students in fulfilling their testing requirements. The section also explores the helpfulness of the information provided about test performance (7c), which includes diagrams, marks, overall assessments, and feedback. This question aims to understand how effectively the performance data is presented and how useful it is for students' self-assessment and improvement.

Additionally, the potential impact of doing tests on SB on students' overall school performance is evaluated (7d), gauging the extent to which students believe these tools contribute to their academic success.

3.1.2.1. Ease of Use (Question 7a)

Ease of performing the test assigned by the teacher varied across the surveyed countries:

- **Cyprus**: 70% of students found it very easy, and 10% found it extremely easy, resulting in 80% having a positive view. 20% were neutral, with no negative views.
- **Greece**: Feedback was mixed, with 36.84% finding it very easy and 29.82% extremely easy. However, 17.54% were moderately easy, and 15.79% had negative views, indicating some difficulty.
- **Ireland**: 35.29% of students found it very easy, and 52.94% found it extremely easy, giving a total of 88.24% positive views. Only 5.88% were neutral or negative.
- **Sweden**: 12.50% found it very easy, and 37.50% found it extremely easy. However, 50% were undecided, showing a high level of uncertainty.

In summary, Cyprus and Ireland reported high satisfaction with ease of use, while Greece and Sweden had more mixed feedback, highlighting areas for potential improvement.

Table 9: 7a. How easy was it to perform the test assigned to you by the teacher?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	10	70	20	0	0	**80,00**	**20,00**	**0,00**
Greece	29,82	36,84	17,54	3,51	12,28	**66,67**	**17,54**	**15,79**
Ireland	52,94	35,29	5,88	5,88	0,00	**88,24**	**5,88**	**5,88**
Sweden	37,50	12,50	50,00	0,00	0,00	**50,00**	**50,00**	**0,00**
Overall	32,61	38,04	18,48	3,26	7,61	**70,65**	**18,48**	**10,87**

Co-funded by the European Union

3.1.2.2. *Impact on Learning (Question 7b)*

Regarding the usefulness of the SB Test Tool to perform tests assigned by the teacher the results were as follows:

- **Cyprus**: 40% found the tool very useful, and 20% found it extremely useful, leading to 60% positive views. 40% were neutral.
- **Greece**: Mixed feedback with 33.33% finding it very useful, 12.28% extremely useful, and 36.84% moderately useful. 17.54% had negative views.
- **Ireland**: 17.65% found it very useful, and 52.94% found it extremely useful, resulting in 70.59% positive views. 17.65% were neutral, and 11.76% had negative views.
- **Sweden**: 37.50% found it very useful, and 12.50% found it extremely useful. However, 25% found it slightly useful, and 12.50% not useful, resulting in a balanced 50% positive and 37.50% negative views.

In summary, while Cyprus and Ireland viewed the tool positively, Greece and Sweden showed more varied responses, suggesting areas for improvement in the tool's usability and effectiveness.

Table 10: 7b. How useful is the Study Buddy Test Tool to perform tests assigned to you by the teacher?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	20	40	40	0	0	**60,00**	**40,00**	**0,00**
Greece	12,28	33,33	36,84	7,02	10,53	**45,61**	**36,84**	**17,54**
Ireland	52,94	17,65	17,65	11,76	0,00	**70,59**	**17,65**	**11,76**
Sweden	12,50	37,50	12,50	25,00	12,50	**50,00**	**12,50**	**37,50**
Overall	20,65	31,52	31,52	8,70	7,61	**52,17**	**31,52**	**16,30**

3.1.2.3. *Overall Satisfaction (Question 7c)*

Students' feedback regarding the helpfulness of the information about test performance (diagram, marks, overall assessment, feedback) is the following:

- **Cyprus**: 50% found the information very helpful, and 30% found it extremely helpful, resulting in 80% positive views. 20% were neutral.
- **Greece**: 31.58% found the information very helpful, and 12.28% found it extremely helpful. However, 31.58% were neutral, and 24.56% had negative views, indicating mixed satisfaction.
- **Ireland**: 35.29% found the information very helpful, and 23.53% found it extremely helpful, resulting in 58.82% positive views. 35.29% were neutral, and 5.88% had negative views.

Co-funded by the European Union

- **Sweden**: 25% found the information very helpful, and no one found it extremely helpful. However, 50% were neutral, and 25% had negative views, showing significant indecision and mixed feedback.

In summary, Cyprus showed high satisfaction with the provided information, while Greece, Ireland, and Sweden had more mixed feedback, highlighting the need for improved clarity and usefulness of performance data.

Table 11: 7c. How helpful did you find the information about your performance on the test (diagram, marks, overall assessment, feedback)?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	30	50	20	0	0	80,00	20,00	0,00
Greece	12,28	31,58	31,58	12,28	12,28	43,86	31,58	24,56
Ireland	23,53	35,29	35,29	5,88	0,00	58,82	35,29	5,88
Sweden	0,00	25,00	50,00	25,00	0,00	25,00	50,00	25,00
Overall	15,22	33,70	32,61	10,87	7,61	48,91	32,61	18,48

3.1.2.4. Conclusions on the Test Taking and Student Dashboard

The feedback on the SB Test Taking and Student Dashboard features varied across different countries, highlighting both strengths and areas for improvement.

In **Cyprus** students generally found the SB Test Tool easy to use and helpful for performing tests. The information provided about test performance was also highly regarded. No significant areas of improvement were identified.

In **Greece** the feedback was mixed. While many students found the tool easy to use and useful, there were significant neutral and negative responses, particularly regarding the usefulness of the test tool and the information about test performance. These findings suggest a need for improvements in usability and clarity of performance data.

In **Ireland**, overall, students had a positive experience with the SB Test Tool, finding it easy to use and useful. However, there were some negative responses, particularly concerning the information about test performance, indicating areas where clarity and detail could be enhanced.

In **Sweden**, the feedback was varied, with a significant number of neutral and negative responses. Students were particularly uncertain about the ease of use and the helpfulness of the performance information, suggesting a need for better user guidance and clearer performance metrics.

Co-funded by the European Union

Summary Table: Areas of Improvement and Action Points

Table 12: Test Taking and Student Dashboard Summary Table

Country	Areas of Improvement	Action Points
Cyprus	None identified	N/A
Greece	Usability, clarity of performance data	Simplify user interface, improve the presentation of performance metrics
Ireland	Clarity and detail of performance information	Enhance feedback details, provide more comprehensive performance data
Sweden	Usability, user guidance, clarity of performance data	Improve user interface, provide better guidance and clearer performance metrics

As an **overall Conclusion**, the SB Test Tool and Student Dashboard features were well-received in Cyprus, with high levels of satisfaction reported. Greece and Sweden had more mixed feedback, highlighting specific areas where the tools could be improved to enhance user experience and the clarity of performance information. Ireland's feedback was generally positive but indicated areas for improvement in the clarity and detail of performance information.

3.1.3. Chat with the Study Buddy

In this section, we examine users' experiences and perceptions regarding the Chat with SB feature. The questions aim to evaluate multiple facets of the chat tool to gather comprehensive feedback. The usefulness of the chat feature in writing is examined first (9a), assessing how beneficial users find text-based interactions with SB. Following this, the ease of using the chat feature in writing is explored (9b), highlighting the user-friendliness and accessibility of the text interface. Additionally, the section investigates the usefulness of voice interactions with SB (9c), providing insights into how helpful users find this alternative mode of communication. The ease of using voice interactions is also assessed (9d), determining how intuitive and accessible the voice interface is for users. The quality of speech recognition (9e) is another critical aspect covered in this section, evaluating the tool's ability to accurately understand and process spoken input. Complementing this, the quality of speech generation (9f) is examined, assessing how well SB can generate clear and coherent spoken responses.

3.1.3.1. Usefulness of Chatting with the Study Buddy in Writing (Question 9a)

Responses on usefulness of chatting with the Study Buddy in writing varied across the surveyed countries:

- **Cyprus**: 60% of students found it very useful, and 30% found it extremely useful, resulting in 90% having a positive view. 10% were neutral, with no negative views.
- **Greece**: Feedback was mixed, with 28.07% finding it very useful, 21.05% extremely useful, and 35.09% moderately useful. 15.79% had negative views.

Co-funded by the European Union

- **Ireland**: 58.82% of students found it very useful, and 17.65% found it extremely useful, giving a total of 76.47% positive views. 11.76% were neutral, and 11.76% had negative views.
- **Sweden**: 62.50% found it very useful, and 12.50% found it moderately useful. However, 25% found it slightly useful, resulting in 62.50% positive views and 25% negative views.

In summary, Cyprus showed the highest satisfaction with the usefulness of chatting in writing, while Greece and Sweden had more mixed feedback, highlighting areas for potential improvement.

Table 13: 9a. How useful was it to chat with Study Buddy in writing?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	30	60	10	0	0	**90,00**	**10,00**	**0,00**
Greece	21,05	28,07	35,09	5,26	10,53	**49,12**	**35,09**	**15,79**
Ireland	17,65	58,82	11,76	11,76	0,00	**76,47**	**11,76**	**11,76**
Sweden	0,00	62,50	12,50	25,00	0,00	**62,50**	**12,50**	**25,00**
Overall	19,57	40,22	26,09	7,61	6,52	**59,78**	**26,09**	**14,13**

3.1.3.2. Ease of Chatting with the Study Buddy in Writing (Question 9b)

- **Cyprus**: 50% found it very easy, and 30% found it extremely easy, resulting in 80% positive views. 20% were neutral, with no negative views.
- **Greece**: 42.11% found it very easy, and 26.32% extremely easy. However, 15.79% were moderately easy, and 15.79% had negative views.
- **Ireland**: 23.53% found it very easy, and 52.94% found it extremely easy, giving a total of 76.47% positive views. 11.76% were neutral, and 11.76% had negative views.
- **Sweden**: 50% found it very easy, and 12.50% found it extremely easy. However, 37.50% were neutral.

In summary, Cyprus and Ireland reported high satisfaction with the ease of use, while Greece had more mixed feedback, indicating areas for improvement in usability.

Table 14: 9b. How easy was it to chat with Study Buddy in writing?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	50	30	20	0	0	**80,00**	**20,00**	**0,00**
Greece	26,32	42,11	15,79	7,02	8,77	**68,42**	**15,79**	**15,79**
Ireland	52,94	23,53	11,76	11,76	0,00	**76,47**	**11,76**	**11,76**

Co-funded by the European Union

Sweden	12,50	50,00	37,50	0,00	0,00	**62,50**	**37,50**	**0,00**
Overall	32,61	38,04	17,39	6,52	5,43	**70,65**	**17,39**	**11,96**

3.1.3.3. Usefulness of Chatting with the Study Buddy Using Voice Interaction (Question 9c)

- **Cyprus**: 40% found it very useful, and 40% found it extremely useful, resulting in 80% positive views. 10% were neutral, and 10% had negative views.
- **Greece**: Mixed feedback with 17.54% finding it very useful, 8.77% extremely useful, and 43.86% moderately useful. 29.82% had negative views.
- **Ireland**: 29.41% found it very useful, and 11.76% found it extremely useful, resulting in 41.18% positive views. 29.41% were neutral, and 29.41% had negative views.
- **Sweden**: 12.50% found it very useful, and 12.50% found it extremely useful. However, 25% found it slightly useful, and 25% not useful, resulting in 25% positive views and 50% negative views.

In summary, Cyprus showed the highest satisfaction with voice interactions, while Greece, Ireland, and Sweden had more mixed feedback, indicating areas for improvement in voice interaction usefulness.

Table 15: 9c. How useful was it to chat with Study Buddy using voice interaction?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	**Positive View (%)**	**Neutral View (%)**	**Negative View (%)**
Cyprus	40	40	10	10	0	**80,00**	**10,00**	**10,00**
Greece	8,77	17,54	43,86	8,77	21,05	**26,32**	**43,86**	**29,82**
Ireland	11,76	29,41	29,41	29,41	0,00	**41,18**	**29,41**	**29,41**
Sweden	12,50	12,50	25,00	25,00	25,00	**25,00**	**25,00**	**50,00**
Overall	13,04	21,74	35,87	14,13	15,22	**34,78**	**35,87**	**29,35**

3.1.3.4. Ease of Chatting with Study Buddy Using Voice Interaction (Question 9d)

- **Cyprus**: 60% found it very easy, and 30% found it extremely easy, resulting in 90% positive views. 10% were neutral, with no negative views.
- **Greece**: 24.56% found it very easy, and 5.26% found it extremely easy. However, 50.88% were neutral, and 19.30% had negative views.
- **Ireland**: 11.76% found it very easy, and 17.65% found it extremely easy, resulting in 29.41% positive views. 29.41% were neutral, and 41.18% had negative views.
- **Sweden**: 25% found it very easy, and none found it extremely easy. However, 25% were neutral, and 50% had negative views.

Co-funded by the European Union

In summary, Cyprus showed the highest satisfaction with the ease of use of voice interactions, while Greece, Ireland, and Sweden had more mixed feedback, indicating areas for improvement in usability.

Table 16: 9d. How easy was it to chat with Study Buddy using voice interaction?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	30	60	10	0	0	**90,00**	**10,00**	**0,00**
Greece	5,26	24,56	50,88	5,26	14,04	**29,82**	**50,88**	**19,30**
Ireland	17,65	11,76	29,41	41,18	0,00	**29,41**	**29,41**	**41,18**
Sweden	0,00	25,00	25,00	37,50	12,50	**25,00**	**25,00**	**50,00**
Overall	9,78	26,09	40,22	14,13	9,78	**35,87**	**40,22**	**23,91**

3.1.3.5. Quality of Speech Recognition (Question 9e)

- **Cyprus**: 50% found the quality very good, and none found it extremely good, resulting in 50% positive views. 40% were neutral, and 10% had negative views.
- **Greece**: 28.07% found the quality very good, and 5.26% extremely good. However, 38.60% were neutral, and 28.07% had negative views.
- **Ireland**: 23.53% found the quality very good, and 35.29% extremely good, resulting in 58.82% positive views. 17.65% were neutral, and 24% had negative views.
- **Sweden**: None found the quality extremely good, and 25% found it very good. However, 12.50% were neutral, and 62.50% had negative views.

In summary, Cyprus and Ireland showed higher satisfaction with the quality of speech recognition, while Greece and Sweden had more mixed feedback, highlighting areas for improvement in speech recognition accuracy.

Table 17: 9e. How good was the quality of speech recognition?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	0	50	40	10	0	**50,00**	**40,00**	**10,00**
Greece	5,26	28,07	38,60	15,79	12,28	**33,33**	**38,60**	**28,07**
Ireland	35,29	23,53	17,65	17,65	5,88	**58,82**	**17,65**	**23,53**
Sweden	25,00	0,00	12,50	37,50	25,00	**25,00**	**12,50**	**62,50**
Overall	11,96	27,17	32,61	17,39	10,87	**39,13**	**32,61**	**28,26**

Co-funded by the European Union

3.1.3.6. *Quality of Speech Generation (Question 9f)*

- **Cyprus**: 40% found the quality very good, and 20% found it extremely good, resulting in 60% positive views. 30% were neutral, and 10% had negative views.
- **Greece**: 26.32% found the quality very good, and 5.26% extremely good. However, 40.35% were neutral, and 28.07% had negative views.
- **Ireland**: 11.76% found the quality very good, and 47.06% extremely good, resulting in 58.82% positive views. 29.41% were neutral, and 11.76% had negative views.
- **Sweden**: None found the quality extremely good, and 37.50% found it very good. However, 37.50% were neutral, and 25% had negative views.

In summary, Cyprus and Ireland showed higher satisfaction with the quality of speech generation, while Greece and Sweden had more mixed feedback, highlighting areas for improvement in speech generation clarity and coherence.

Table 188: 9f. How good was the quality of speech generation?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	20	40	30	10	0	**60,00**	**30,00**	**10,00**
Greece	5,26	26,32	40,35	17,54	10,53	**31,58**	**40,35**	**28,07**
Ireland	47,06	11,76	29,41	5,88	5,88	**58,82**	**29,41**	**11,76**
Sweden	0,00	37,50	37,50	25,00	0,00	**37,50**	**37,50**	**25,00**
Overall	14,13	26,09	36,96	15,22	7,61	**40,22**	**36,96**	**22,83**

3.1.3.7. *Conclusions on Chat with Study Buddy*

The feedback on the Chat with Study Buddy feature varied across different countries, highlighting both strengths and areas for improvement.

In **Cyprus**, students generally found the chat features (both text and voice) useful and easy to use. The quality of speech recognition and generation was also well-regarded. No significant areas of improvement were identified.

In **Greece**, the feedback was mixed. While many students found the chat features useful and easy to use, there were significant neutral and negative responses, particularly regarding the usefulness and ease of voice interaction, and the quality of speech recognition and generation. These findings suggest a need for improvements in voice interaction and speech processing.

In **Ireland**, overall, students had a positive experience with the chat features, finding them useful and easy to use. However, there were some negative responses, particularly

Co-funded by the European Union

concerning the quality of speech recognition and generation, indicating areas where clarity and accuracy could be enhanced.

In **Sweden**: The feedback was varied, with a significant number of neutral and negative responses. Students were particularly uncertain about the usefulness and ease of voice interaction, and the quality of speech recognition and generation, suggesting a need for better user guidance and improved speech processing accuracy.

Summary Table: Areas of Improvement and Action Points

Table 19: Chat with SB - Summary Table

Country	Areas of Improvement	Action Points
Cyprus	None identified	N/A
Greece	Voice interaction, speech recognition, speech generation	Enhance voice interaction usability, improve speech recognition and generation accuracy
Ireland	Speech recognition, speech generation	Improve clarity and accuracy of speech recognition and generation
Sweden	Voice interaction, speech recognition, speech generation	Improve user guidance, enhance speech processing accuracy

As an **overall conclusion**, the Chat with Study Buddy feature was well-received in Cyprus, with high levels of satisfaction reported. Greece and Sweden had more mixed feedback, highlighting specific areas where the tools could be improved to enhance user experience and the accuracy of speech processing. Ireland's feedback was generally positive but indicated areas for improvement in the quality of speech recognition and generation.

3.1.4. Chat History

In this section, we explore users' experiences with the Chat History feature of SB. The following questions aim to gather insights on various aspects of this feature. Firstly, users were asked about the accuracy of the short titles for their previous chats (11a). This question seeks to understand if the titles effectively summarize the content of past interactions, making it easier for users to identify and revisit important conversations. Secondly, the ease of navigating through the Chat History was assessed (11b), providing valuable information on the user-friendliness and accessibility of this feature. Additionally, after reviewing the chats performed with the SB in chat history, the participants were asked to evaluate the overall accuracy of the information provided in SB's responses (11c), to examine the reliability and precision of the answers. Lastly, the correctness of the language used in SB's responses was examined (11d), highlighting the importance of clear and grammatically accurate communication. Collectively, these questions aim to identify strengths and areas for improvement regarding the Chat History feature, as well as regarding the overall quality of AI's responses in terms of accuracy and language.

Co-funded by the European Union

3.1.4.1. *Accuracy of Short Titles of Previous Chats (Question 11a)*

Accuracy of short titles of previous chats varied across the surveyed countries:

- **Cyprus**: 90% of students found the titles very accurate, and 10% found them extremely accurate, resulting in 100% having a positive view with no negative or neutral views.
- **Greece**: Feedback was mixed, with 29.82% finding them very accurate, 10.53% extremely accurate, and 50.88% moderately accurate. 8.77% had negative views.
- **Ireland**: 35.29% of students found the titles very accurate, and 41.18% found them extremely accurate, giving a total of 76.47% positive views. 5.88% were neutral, and 17.65% had negative views.
- **Sweden**: 50% found the titles very accurate, and 25% found them extremely accurate. 25% were neutral with no negative views.

Table 2020: 11a. How accurate are the short titles of your previous chats?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	10	90	0	0	0	**100,00**	**0,00**	**0,00**
Greece	10,53	29,82	50,88	0,00	8,77	**40,35**	**50,88**	**8,77**
Ireland	41,18	35,29	5,88	17,65	0,00	**76,47**	**5,88**	**17,65**
Sweden	25,00	50,00	25,00	0,00	0,00	**75,00**	**25,00**	**0,00**
Overall	17,39	39,13	34,78	3,26	5,43	**56,52**	**34,78**	**8,70**

3.1.4.2. *Ease of Going Through the Chat History (Question 11b)*

- **Cyprus**: 70% found it extremely easy, and 30% found it very easy, resulting in 100% positive views with no neutral or negative views.
- **Greece**: 33.33% found it very easy, and 10.53% found it extremely easy. However, 47.37% were moderately easy, and 8.77% had negative views.
- **Ireland**: 23.53% found it very easy, and 52.94% found it extremely easy, giving a total of 76.47% positive views. 5.88% were neutral, and 17.65% had negative views.
- **Sweden**: 25% found it very easy, and 62.50% found it extremely easy. 12.50% were neutral with no negative views.

Table 211: 11b. How easy was it to go through the Chat History?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	70	30	0	0	0	**100,00**	**0,00**	**0,00**
Greece	10,53	33,33	47,37	0,00	8,77	**43,86**	**47,37**	**8,77**

Co-funded by the European Union

Ireland	52,94	23,53	5,88	17,65	0,00	**76,47**	**5,88**	**17,65**
Sweden	62,50	25,00	12,50	0,00	0,00	**87,50**	**12,50**	**0,00**
Overall	29,35	30,43	31,52	3,26	5,43	**59,78**	**31,52**	**8,70**

3.1.4.3. Overall Accuracy of Information Provided in Study Buddy's Responses (Question 11c)

Students feedback on the overall accuracy of the SB responses is presented below:

- **Cyprus**: 50% found the information very accurate, and 20% found it extremely accurate, resulting in 70% positive views. 30% were neutral with no negative views.
- **Greece**: 38.60% found the information very accurate, and 17.54% extremely accurate. However, 35.09% were neutral, and 8.77% had negative views.
- **Ireland**: 17.65% found the information very accurate, and 58.82% found it extremely accurate, giving a total of 76.47% positive views. There were no neutral views, and 23.53% had negative views.
- **Sweden**: 62.50% found the information very accurate, and 12.50% found it extremely accurate. 12.50% were neutral, and 12.50% had negative views.

In summary, Cyprus and Ireland showed higher satisfaction with the overall accuracy of information provided, while Greece and Sweden had more mixed feedback, highlighting areas for improvement in accuracy.

Table 222: 11c. Overall, how accurate was the information provided in the Study Buddy's responses?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	20	50	30	0	0	**70,00**	**30,00**	**0,00**
Greece	17,54	38,60	35,09	0,00	8,77	**56,14**	**35,09**	**8,77**
Ireland	58,82	17,65	0,00	17,65	5,88	**76,47**	**0,00**	**23,53**
Sweden	12,50	62,50	12,50	12,50	0,00	**75,00**	**12,50**	**12,50**
Overall	25,00	38,04	26,09	4,35	6,52	**63,04**	**26,09**	**10,87**

3.1.4.4. Overall Correctness of the Language Used in Study Buddy's Responses (Question 11d)

Regarding the correctness of language, students gave the following feedback:

- **Cyprus**: 40% found the language very correct, and 50% found it extremely correct, resulting in 90% positive views. 10% were neutral with no negative views.
- **Greece**: 35.09% found the language very correct, and 14.04% extremely correct. However, 40.35% were neutral, and 10.53% had negative views.

Co-funded by the European Union

- **Ireland**: 29.41% found the language very correct, and 52.94% found it extremely correct, giving a total of 82.35% positive views. 5.88% were neutral, and 11.76% had negative views.
- **Sweden**: 50% found the language very correct, and 12.50% found it extremely correct. 37.50% were neutral with no negative views.

In summary, Cyprus and Ireland showed higher satisfaction with the correctness of the language used, while Greece and Sweden had more mixed feedback, highlighting areas for improvement in language accuracy.

Table 233: 11d. Overall, how correct was the language used in the Study Buddy's responses?

Country	Extremely (%)	Very (%)	Moderately (%)	Slightly (%)	Not at all (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	50	40	10	0	0	**90,00**	**10,00**	**0,00**
Greece	14,04	35,09	40,35	0,00	10,53	**49,12**	**40,35**	**10,53**
Ireland	52,94	29,41	5,88	11,76	0,00	**82,35**	**5,88**	**11,76**
Sweden	12,50	50,00	37,50	0,00	0,00	**62,50**	**37,50**	**0,00**
Overall	25,00	35,87	30,43	2,17	6,52	**60,87**	**30,43**	**8,70**

3.1.4.5. Conclusions on Chat History and Evaluation of Quality of Study Buddy's Responses

The feedback on the Chat History and evaluation of the quality of Study Buddy's responses varied across different countries, highlighting both strengths and areas for improvement.

In **Cyprus** Students generally found the Chat History accurate, easy to navigate, and the language used in responses correct. No significant areas of improvement were identified.

In **Greece** the feedback was mixed. While many students found the features accurate and easy to use, there were significant neutral and negative responses, particularly regarding the accuracy of information and the correctness of language used. These findings suggest a need for improvements in accuracy and language precision.

In **Ireland** students had a positive experience with the Chat History feature, finding it accurate and easy to use. However, there were some negative responses, particularly concerning the correctness of language, indicating areas where clarity and precision could be enhanced.

In **Sweden** the feedback was varied, with a significant number of neutral and negative responses. Students were particularly uncertain about the ease of navigating the Chat History and the correctness of the language used, suggesting a need for better user guidance and improved language accuracy.

Co-funded by the European Union

Summary Table: Areas of Improvement and Action Points

Table 24: Chat History - Summary Table

Country	Areas of Improvement	Action Points
Cyprus	None identified	N/A
Greece	Accuracy of information, correctness of language	Improve information accuracy, enhance language precision
Ireland	Correctness of language	Improve clarity and precision of language
Sweden	Ease of navigation, correctness of language	Enhance user guidance, improve language accuracy

As an overall conclusion, the Chat History feature and the quality of Study Buddy's responses were well-received in Cyprus, with high levels of satisfaction reported. Greece and Sweden had more mixed feedback, highlighting specific areas where the tools could be improved to enhance user experience and the accuracy of information and language used. Ireland's feedback was generally positive but indicated areas for improvement in the correctness of language used in responses.

3.1.5. Usability, functionality, and impact on learning

This section gathers detailed feedback on the SB tool, focusing on its usability, functionality, and impact on learning. The questions address various aspects of user experience. Users are asked about their satisfaction with the SB graphical user interface (**Q13a**), evaluating its appeal and user-friendliness. The frequency of use (**Q13b**) is explored to understand how often users would like to use SB, indicating the tool's effectiveness and user satisfaction. Users assess whether they find the SB tools unnecessarily complex (**Q13c**), as complexity can be a significant barrier to effective usage. The ease of navigation (**Q13d**) is evaluated to ensure that users can quickly find and utilize features. Additionally, the overall ease of use (**Q13e**) is examined, highlighting how intuitive and accessible the tool is.

Further questions investigate the clarity of interactions (**Q13f**), assessing if users find their interactions with SB clear and understandable. The usefulness of SB in daily homework (**Q13g**) and its potential to empower users as efficient users of AI (Q13h) are also explored, providing insights into the tool's practical benefits and relevance in academic routines. Users are asked if SB makes learning more interesting (**Q13i**) and if it improves their engagement in learning activities (**Q13j**). These questions aim to determine the tool's impact on academic performance and knowledge acquisition.

The survey also examines if SB helps users understand their study subjects better (**Q13k**) and improves the quality of their everyday study (**Q13l**). Users' access to necessary resources such as a PC, workstation, internet connection, and data to use SB effectively (**Q13m**) is also queried, identifying any external barriers to usage. Additionally, the survey investigates users' autonomy in terms of knowledge and skills to use the SB (**Q13n**), asking if they can use SB without external support (**Q13o**). Users are also asked about their intention to use the Study Buddy in the future (Q13p) and whether they believe using the Study Buddy will improve their performance at school (**Q13q**).

Co-funded by the European Union

3.1.5.1. *Graphical User Interface of the Study Buddy (Question 13a)*

The feedback on the graphical user interface (GUI) shows generally positive perceptions across all surveyed countries, with varying degrees of satisfaction.

Table 25: 13a. I like the Graphical user interface of Study Buddy

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	40	50	10	0	0	**90,00**	**10,00**	**0,00**
Greece	8,77	54,39	29,82	1,75	5,26	**63,16**	**29,82**	**7,02**
Ireland	82,35	17,65	0,00	0,00	0,00	**100,00**	**0,00**	**0,00**
Sweden	12,50	75,00	0,00	12,50	0,00	**87,50**	**0,00**	**12,50**
Overall	26,09	48,91	19,57	2,17	3,26	**75,00**	**19,57**	**5,43**

- **Cyprus**: 90% of users have a positive view, indicating high satisfaction with the GUI.
- **Greece**: 63.16% positive view with 29.82% undecided, suggesting room for improvement.
- **Ireland**: 100% positive view, reflecting unanimous satisfaction.
- **Sweden**: 87.50% positive view, with a small segment (12.50%) expressing dissatisfaction.

3.1.5.2. *Frequency of Potential Use of Study Buddy (Question 13b)*

The willingness to use SB frequently varies among countries, reflecting different levels of user engagement.

Table 265: 13b. I think that I would like to use the Study Buddy frequently.

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	70	10	20	0	0	**80,00**	**20,00**	**0,00**
Greece	14,04	29,82	40,35	5,26	10,53	**43,86**	**40,35**	**15,79**
Ireland	58,82	35,29	5,88	0,00	0,00	**94,12**	**5,88**	**0,00**
Sweden	12,50	25,00	37,50	12,50	12,50	**37,50**	**37,50**	**25,00**
Overall	28,26	28,26	31,52	4,35	7,61	**56,52**	**31,52**	**11,96**

- **Cyprus**: 80% of users express a desire to use SB frequently.
- **Greece**: 43.86% positive view with 40.35% undecided, indicating a need for better engagement.
- **Ireland**: 94.12% positive view, suggesting strong engagement.
- **Sweden**: 37.50% positive view with a high neutral view (37.50%) and 25% negative, reflecting mixed feelings.

Co-funded by the European Union

3.1.5.3. *Complexity of Study Buddy Tools (Question 13c)*

The perception of complexity varies, indicating areas where the tools could be simplified.

Table 276: 13c. The Study Buddy tools are unnecessarily complex.

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	10	0	0	90	0	**10,00**	**0,00**	**90,00**
Greece	7,02	10,53	26,32	40,35	15,79	**17,54**	**26,32**	**56,14**
Ireland	17,65	17,65	0,00	41,18	23,53	**35,29**	**0,00**	**64,71**
Sweden	0,00	12,50	12,50	62,50	12,50	**12,50**	**12,50**	**75,00**
Overall	8,70	10,87	17,39	47,83	15,22	**19,57**	**17,39**	**63,04**

- **Cyprus**: 90% of users find the tools unnecessarily complex.
- **Greece**: 56.14% negative view, indicating significant complexity issues.
- **Ireland**: 64.71% negative view, suggesting room for simplification.
- **Sweden**: 75% negative view, highlighting a need for improvement in tool complexity.

3.1.5.4. *Ease of Navigation of the Study Buddy (Question 13d)*

Users generally find SB easy to navigate, though some areas could be improved.

Table 287: 13d. I think the Study Buddy is easy to navigate.

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	40	60	0	0	0	**100,00**	**0,00**	**0,00**
Greece	17,54	49,12	22,81	3,51	7,02	**66,67**	**22,81**	**10,53**
Ireland	52,94	47,06	0,00	0,00	0,00	**100,00**	**0,00**	**0,00**
Sweden	37,50	50,00	0,00	12,50	0,00	**87,50**	**0,00**	**12,50**
Overall	28,26	50,00	14,13	3,26	4,35	**78,26**	**14,13**	**7,61**

- **Cyprus and Ireland**: 100% positive view, indicating excellent navigability.
- **Greece**: 66.67% positive view, but 22.81% undecided and 10.53% negative, showing room for improvement.
- **Sweden**: 87.50% positive view, with some users (12.50%) experiencing difficulties.

3.1.5.5. *Ease of Use of Study Buddy (Question 13e)*

The tool's ease of use is generally well-received, though some users face challenges.

Co-funded by the European Union

Table 298: 13e. I find the Study Buddy easy to use.

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	40	50	0	0	10	**90,00**	**0,00**	**10,00**
Greece	29,82	42,11	17,54	5,26	5,26	**71,93**	**17,54**	**10,53**
Ireland	52,94	47,06	0,00	0,00	0,00	**100,00**	**0,00**	**0,00**
Sweden	37,50	50,00	0,00	12,50	0,00	**87,50**	**0,00**	**12,50**
Overall	35,87	44,57	10,87	4,35	4,35	**80,43**	**10,87**	**8,70**

- **Cyprus and Ireland**: High satisfaction with ease of use.
- **Greece**: 71.93% positive view, but 17.54% undecided and 10.53% negative, indicating some usability issues.
- **Sweden**: 87.50% positive view, with a small segment (12.50%) facing usability challenges.

3.1.5.6. Clearness of Interactions with the Study Buddy (Question 13f)

Feedback on interaction clarity is generally positive, but with room for improvement.

Table 309: 13f. My interaction with the Study Buddy was clear and understandable.

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	20	60	20	0	0	**80,00**	**20,00**	**0,00**
Greece	12,28	42,11	33,33	3,51	8,77	**54,39**	**33,33**	**12,28**
Ireland	47,06	52,94	0,00	0,00	0,00	**100,00**	**0,00**	**0,00**
Sweden	25,00	37,50	0,00	37,50	0,00	**62,50**	**0,00**	**37,50**
Overall	20,65	45,65	22,83	5,43	5,43	**66,30**	**22,83**	**10,87**

- **Cyprus and Ireland**: High satisfaction with interaction clarity.
- **Greece**: 54.39% positive view with 33.33% neutral and 12.28% negative, indicating some clarity issues.
- **Sweden**: Mixed responses with 62.50% positive and 37.50% negative, highlighting significant clarity issues.

3.1.5.7. Daily Use of Study Buddy (Question 13g)

User opinions on SB's utility in daily homework vary, indicating areas for improvement.

Table30: 13g. I think that the Study Buddy will be useful in my daily homework.

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)

Co-funded by the European Union

Cyprus	40	30	10	20	0	**70,00**	**10,00**	**20,00**
Greece	17,54	26,32	33,33	7,02	15,79	**43,86**	**33,33**	**22,81**
Ireland	70,59	17,65	11,76	0,00	0,00	**88,24**	**11,76**	**0,00**
Sweden	12,50	25,00	50,00	0,00	12,50	**37,50**	**50,00**	**12,50**
Overall	29,35	25,00	28,26	6,52	10,87	**54,35**	**28,26**	**17,39**

- **Cyprus**: 70% positive view but 20% negative, indicating areas for improvement.
- **Greece**: 43.86% positive with 33.33% neutral and 22.81% negative, reflecting mixed feelings.
- **Ireland**: 88.24% positive view, indicating high satisfaction.
- **Sweden**: 37.50% positive with 50% neutral, suggesting users are still evaluating the tool's benefits.

3.1.5.8. Empowering Users of the Study Buddy (Question 13h)

Users' perceptions of SB's ability to empower them as efficient AI users vary widely.

Table 31: 13h. I think that Study Buddy can empower me as an efficient user of Artificial Intelligence.

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	30	50	20	0	0	**80,00**	**20,00**	**0,00**
Greece	15,79	29,82	35,09	8,77	10,53	**45,61**	**35,09**	**19,30**
Ireland	41,18	58,82	0,00	0,00	0,00	**100,00**	**0,00**	**0,00**
Sweden	0,00	37,50	62,50	0,00	0,00	**37,50**	**62,50**	**0,00**
Overall	20,65	38,04	29,35	5,43	6,52	**58,70**	**29,35**	**11,96**

- **Cyprus**: 80% positive view, indicating strong confidence in SB's empowering potential.
- **Greece**: 45.61% positive view with 35.09% undecided and 19.30% negative, reflecting mixed perceptions.
- **Ireland**: 100% positive view, showing unanimous confidence in SB.
- **Sweden**: 37.50% positive view with a high neutral view (62.50%), suggesting users need more time to evaluate SB's effectiveness.

3.1.5.9. Learning Experience with the Study Buddy (Question 13i)

Users generally find the learning experience with SB interesting, though some are undecided.

Table 32: 13i. I find the experience of learning with the Study Buddy interesting.

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	60	30	10	0	0	**90,00**	**10,00**	**0,00**
Greece	19,30	33,33	26,32	10,53	10,53	**52,63**	**26,32**	**21,05**

Co-funded by the European Union

Ireland	41,18	58,82	0,00	0,00	0,00	**100,00**	**0,00**	**0,00**
Sweden	0,00	75,00	0,00	25,00	0,00	**75,00**	**0,00**	**25,00**
Overall	26,09	41,30	17,39	8,70	6,52	**67,39**	**17,39**	**15,22**

- **Cyprus and Ireland**: High satisfaction with the learning experience.
- **Greece**: 52.63% positive with 26.32% neutral and 21.05% negative, indicating mixed feelings.
- **Sweden**: 75% positive view, but 25% negative, reflecting a need for improvement in making learning more engaging.

3.1.5.10. Engagement in Learning through the Study Buddy (Question 13j)

Users' perceptions of SB's impact on their engagement in learning vary widely.

Table 33: 13j. I think that using the Study Buddy improves my engagement in learning.

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	30	60	10	0	0	**90,00**	**10,00**	**0,00**
Greece	12,28	21,05	33,33	21,05	12,28	**33,33**	**33,33**	**33,33**
Ireland	47,06	47,06	5,88	0,00	0,00	**94,12**	**5,88**	**0,00**
Sweden	0,00	37,50	62,50	0,00	0,00	**37,50**	**62,50**	**0,00**
Overall	19,57	31,52	28,26	13,04	7,61	**51,09**	**28,26**	**20,65**

- **Cyprus**: 90% positive view, indicating strong perceived engagement.
- **Greece**: 33.33% positive, 33.33% neutral, and 33.33% negative, showing divided opinions.
- **Ireland**: 94.12% positive view, indicating high perceived engagement.
- **Sweden**: 37.50% positive with 62.50% neutral, suggesting users are still evaluating the tool's impact on engagement.

3.1.5.11. Improving Subject Knowledge through the Study Buddy (Question 13k)

Most users find SB helpful in increasing their subject knowledge, with some areas for improvement.

Table 34: 13k. Using the Study Buddy enables me to know more about the subject that I am studying.

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	30	70	0	0	0	**100,00**	**0,00**	**0,00**
Greece	10,53	56,14	19,30	5,26	8,77	**66,67**	**19,30**	**14,04**
Ireland	52,94	41,18	5,88	0,00	0,00	**94,12**	**5,88**	**0,00**
Sweden	37,50	37,50	12,50	0,00	12,50	**75,00**	**12,50**	**12,50**
Overall	22,83	53,26	14,13	3,26	6,52	**76,09**	**14,13**	**9,78**

Co-funded by the European Union

- **Cyprus**: 100% positive view, indicating unanimous satisfaction.
- **Greece**: 66.67% positive with 19.30% neutral and 14.04% negative, indicating mixed feelings.
- **Ireland**: 94.12% positive view, showing strong perceived knowledge improvement.
- **Sweden**: 75% positive view, with some neutral (12.50%) and negative (12.50%) perceptions.

3.1.5.12. Quality of Everyday Studying with the Study Buddy (Question 13l)

Users' perceptions of SB's impact on their daily study quality vary, indicating potential for enhancement.

Table 35: 13l. By using the Study Buddy, the quality of my everyday study will be improved.

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	20	50	30	0	0	**70,00**	**30,00**	**0,00**
Greece	7,02	26,32	40,35	10,53	15,79	**33,33**	**40,35**	**26,32**
Ireland	52,94	41,18	5,88	0,00	0,00	**94,12**	**5,88**	**0,00**
Sweden	12,50	37,50	37,50	0,00	12,50	**50,00**	**37,50**	**12,50**
Overall	17,39	32,61	32,61	6,52	10,87	**50,00**	**32,61**	**17,39**

- **Cyprus**: 70% positive view with 30% neutral, indicating satisfaction with room for improvement.
- **Greece**: 33.33% positive, 40.35% neutral, and 26.32% negative, reflecting mixed opinions.
- **Ireland**: 94.12% positive view, indicating high perceived improvement in study quality.
- **Sweden**: 50% positive with 37.50% neutral and 12.50% negative, showing potential for enhancement.

3.1.5.13. Availability of Resources to Use the Study Buddy (Question 13m)

Users' confidence in having the necessary resources to use SB varies, highlighting areas for improvement.

Table 36: 13m. I have the necessary resources (PC, workstation, Internet Connection, Data) to use the Study Buddy.

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	40	40	20	0	0	**80,00**	**20,00**	**0,00**
Greece	28,07	35,09	29,82	0,00	7,02	**63,16**	**29,82**	**7,02**
Ireland	41,18	29,41	11,76	11,76	5,88	**70,59**	**11,76**	**17,65**
Sweden	62,50	25,00	0,00	12,50	0,00	**87,50**	**0,00**	**12,50**

Co-funded by the European Union

Overall	34,78	33,70	22,83	3,26	5,43	**68,48**	**22,83**	**8,70**

- **Cyprus**: 80% positive view, indicating confidence in resource availability.
- **Greece**: 63.16% positive, 29.82% neutral, and 7.02% negative, reflecting mixed confidence.
- **Ireland**: 70.59% positive with 17.65% negative, indicating room for improvement.
- **Sweden**: 87.50% positive view, showing strong confidence in resource availability.

3.1.5.14. Existing Knowledge to Use the Study Buddy (Question 13n)

Users' confidence in having the necessary knowledge to use SB varies, highlighting areas for additional training or support.

Table 37: 13n. I have the necessary knowledge to use the Study Buddy.

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	30	60	10	0	0	**90,00**	**10,00**	**0,00**
Greece	17,54	43,86	19,30	10,53	8,77	**61,40**	**19,30**	**19,30**
Ireland	58,82	41,18	0,00	0,00	0,00	**100,00**	**0,00**	**0,00**
Sweden	62,50	25,00	0,00	12,50	0,00	**87,50**	**0,00**	**12,50**
Overall	30,43	43,48	13,04	7,61	5,43	**73,91**	**13,04**	**13,04**

- **Cyprus**: 90% positive view, indicating high confidence in knowledge.
- **Greece**: 61.40% positive with 19.30% neutral and 19.30% negative, showing mixed confidence.
- **Ireland**: 100% positive view, reflecting unanimous confidence in knowledge.
- **Sweden**: 87.50% positive view, indicating strong confidence in knowledge.

3.1.5.15. Need for Support While Using the Study Buddy (Question 13o)

Users' perceptions of their ability to use SB independently vary, highlighting areas for additional support.

Table 38: 13o. I could use the Study Buddy, even if there was no one around to support me.

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	60	20	20	0	0	**80,00**	**20,00**	**0,00**
Greece	10,53	40,35	36,84	1,75	10,53	**50,88**	**36,84**	**12,28**
Ireland	58,82	41,18	0,00	0,00	0,00	**100,00**	**0,00**	**0,00**
Sweden	62,50	25,00	0,00	12,50	0,00	**87,50**	**0,00**	**12,50**
Overall	29,35	36,96	25,00	2,17	6,52	**66,30**	**25,00**	**8,70**

- **Cyprus**: 80% positive view, indicating strong perceived independence.

Co-funded by the European Union

- **Greece**: 50.88% positive, 36.84% neutral, and 12.28% negative, reflecting mixed perceptions.
- **Ireland**: 100% positive view, indicating unanimous confidence in using SB independently.
- **Sweden**: 87.50% positive view, showing strong perceived independence.

3.1.5.16. Intention to Use the Study Buddy in the Future (Question 13p)

Users' intentions to continue using SB vary, reflecting different levels of commitment.

Table 39: 13p. I intend to use the Study Buddy in the future.

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	60	40	0	0	0	**100,00**	**0,00**	**0,00**
Greece	7,02	28,07	45,61	8,77	10,53	**35,09**	**45,61**	**19,30**
Ireland	64,71	35,29	0,00	0,00	0,00	**100,00**	**0,00**	**0,00**
Sweden	12,50	25,00	62,50	0,00	0,00	**37,50**	**62,50**	**0,00**
Overall	23,91	30,43	33,70	5,43	6,52	**54,35**	**33,70**	**11,96**

- **Cyprus and Ireland**: 100% positive view, indicating strong commitment to future use.
- **Greece**: 35.09% positive with 45.61% neutral and 19.30% negative, reflecting mixed intentions.
- **Sweden**: 37.50% positive with 62.50% neutral, suggesting users are still deciding on future use.

3.1.5.17. Estimation of Performance at School by Using the Study Buddy (Question 13q)

Users' perceptions of SB's impact on their school performance vary, indicating potential for improvement.

Table 40: 13q. By using the Study Buddy, my performance at school will be improved

Country	Strongly Agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly Disagree (%)	Positive View (%)	Neutral View (%)	Negative View (%)
Cyprus	30	40	30	0	0	**70,00**	**30,00**	**0,00**
Greece	7,02	19,30	42,11	17,54	14,04	**26,32**	**42,11**	**31,58**
Ireland	52,94	41,18	5,88	0,00	0,00	**94,12**	**5,88**	**0,00**
Sweden	12,50	37,50	37,50	0,00	12,50	**50,00**	**37,50**	**12,50**
Overall	18,48	27,17	33,70	10,87	9,78	**45,65**	**33,70**	**20,65**

- **Cyprus**: 70% positive view, indicating strong perceived improvement in performance.

Co-funded by the European Union

- **Greece**: 26.32% positive, 42.11% neutral, and 31.58% negative, reflecting mixed perceptions.
- **Ireland**: 94.12% positive view, showing strong perceived improvement in performance.
- **Sweden**: 50% positive with 37.50% neutral and 12.50% negative, indicating room for improvement.

3.1.5.18. Summary of Results of overall usability, functionality and impact of SB

The analysis of the usability, functionality, and impact of the SB tool across various countries—Cyprus, Greece, Ireland, and Sweden—provides significant insights into user experiences and perceptions. The following table summarizes key findings drawn from the data:

	Positive Feedback	Negative/Neutral Feedback	Countries Needing Improvement
Overall User Satisfaction and Interface Appeal			
Graphical User Interface	Users across all countries generally found the SB interface appealing and user-friendly. The highest satisfaction was noted in Ireland (100% positive) and Cyprus (90% positive).	Some negative and neutral views in Greece and Sweden.	Greece, Sweden
Ease of Navigation and Use	Most users found SB easy to navigate and use, particularly in Cyprus and Ireland where satisfaction was at 100%.	Greece and Sweden indicated areas for improvement in making the tool more intuitive.	Greece, Sweden
Frequency of Use and Complexity			
Frequency of Use	Users in Ireland (94%) and Cyprus (80%) were more inclined to use SB frequently.	Greece and Sweden had mixed responses with notable neutral and negative views.	Greece, Sweden
Complexity	Majority of users in Cyprus and Sweden did not find SB tools unnecessarily complex.	Significant number of users in Greece and Ireland felt the tools were complex.	Greece, Ireland
Learning and Engagement			
Learning Experience	Cyprus (90%) and Ireland (100%) reported high satisfaction with their learning experiences using SB.	Greece and Sweden had mixed views with a notable percentage of users undecided or negative.	Greece, Sweden
Engagement in Learning	Ireland (94%) and Cyprus (90%) had high positive views on SB improving learning engagement.	Greece and Sweden had a more balanced view with a significant percentage of undecided and negative responses.	Greece, Sweden
Academic Support and Knowledge Enhancement			

Co-funded by the European Union

	Positive Feedback	Negative/Neutral Feedback	Countries Needing Improvement
Improving Subject Knowledge	Most users agreed that SB helped them understand their subjects better, with the highest positive responses from Cyprus (100%) and Ireland (94%).	Greece and Sweden indicated some areas for improvement.	Greece, Sweden
Daily Homework Utility	The utility of SB in daily homework was viewed positively by users in Cyprus (70%) and Ireland (88%).	Greece and Sweden showed more mixed responses with significant neutral and negative views.	Greece, Sweden
Impact on School Performance	Users in Ireland (94%) and Cyprus (70%) believed SB would improve their school performance.	Greece and Sweden had more mixed views, with many users undecided or negative.	Greece, Sweden
Resources and Knowledge			
Availability of Resources	Most users in Cyprus, Sweden, and Ireland felt they had the necessary resources to use SB effectively.	Notable percentage in Greece and Ireland were unsure or disagreed.	Greece, Ireland
Existing Knowledge	Users in Cyprus and Ireland felt confident in their knowledge to use SB.	Greece and Sweden showed mixed confidence levels.	Greece, Sweden
Autonomy and Future Intentions			
Need for Support	High levels of perceived autonomy were reported in Cyprus and Ireland (both 100%).	Greece and Sweden had more mixed responses.	Greece, Sweden
Future Use	Users in Cyprus and Ireland expressed strong intentions to continue using SB (both 100%).	Greece and Sweden had significant undecided responses.	Greece, Sweden

3.1.6. Conclusions of quantitative analysis

The SB tool has been generally well-received across different countries, with high satisfaction noted particularly in Cyprus and Ireland. Key strengths include the user-friendly interface, ease of navigation, and perceived improvement in learning and engagement. However, areas for improvement have been identified, especially in Greece and Sweden, where users reported mixed feelings about the complexity of the tools, engagement levels, and the impact on academic performance. To enhance the overall user experience, the following action points emerged:

- Simplify tools and provide better user guidance and training to address perceived complexity, to make learning more interesting and improve daily homework utility.
- Ensure resource accessibility and provide additional training to boost user confidence.
- Demonstrate the long-term benefits and impact of SB on academic performance to encourage continued use.

Co-funded by the European Union

By addressing these areas, the Study Buddy tool can further improve its usability, functionality, and overall impact on learning for a wider range of users.

3.2. Qualitative Analysis

This section presents the feedback collected from students through open-ended questions in the questionnaire. These questions aimed to gather users' suggestions, recommendations for improvement, and insights into SB's tools and features perceived usability. The answers to these open-ended questions complement the quantitative analysis presented in section 3.1. Subsections follow the structure of the 3.1 subsections, as each questionnaire section included both closed and open-ended questions.

3.3.1. Experience Using the SB Tools

The study of the SB tools across Greece, Cyprus, Ireland, and Sweden reveals insightful feedback from students through open-ended questions. These insights cover various aspects of the tools, including their usefulness, ease of use, educational impact, and areas for improvement. Overall, students found the SB tools helpful and easy to use, indicating that these tools are well-designed and effective in supporting learning. However, there are suggestions for improvement to make the tools more universally impactful. Continuous user feedback and iterative development are essential to enhance their effectiveness and meet the diverse needs of all students.

Positive Aspects:

- Students appreciated the current functionalities, with some describing them as "perfect".
- The tools were found helpful and easy to use by many students.
- The text summarization feature was well-received, especially by students with special needs.

Students' Suggestions for Improvement:

- Students expressed the desire for more books and tools for learning foreign languages.
- Practical tools such as self-control features to manage study time and a tool to solve math problems with explanations were suggested.
- Enhancing AI's natural language processing and system performance to handle misspelled words and off-topic answers.
- More concise and simple answers were requested to make the tool more accessible.

Students' Critical Feedback:

- Issues with AI's difficulty with misspelled words and occasional off-topic answers.
- Discrepancies with book or internet information were noted.
- Students noted that the tool failed to address specific questions or provided irrelevant information.

Co-funded by
the European Union

3.3.2. Test and Student Dashboard

The study of the Test and Student Dashboard features of SB tools across Greece, Cyprus, Sweden, and Ireland reveals generally positive feedback from students. The open-ended responses from students in these countries provide valuable insights into the usability, effectiveness, and areas for improvement of these tools.

Positive Aspects:

- Students across the surveyed countries found the Test and Dashboard features easy to use, impactful on their learning, and satisfactory overall.
- In Greece, comments like "It's a nice app, very usable" and "Everything is great. It's a pleasure" highlight the positive aspects.
- In Cyprus, the tools were described as useful, wonderful, and very helpful.
- In Ireland, students expressed high levels of satisfaction with comments like "The tests are ready immediately, and we can complete them quickly without having to wait forever when we're finished—this is fantastic!"

Students' Suggestions for Improvement:

- Desire for more diverse resources and questions: "Possibility of more sources on different subjects" and "I'd like a few more questions".
- Improving the UI to make it more intuitive or visually appealing: "As I mentioned before, make the UI better".
- Enabling teachers to send results to parents and allowing parents to provide feedback.

Students' Critical Feedback:

- Some students in Greece felt a disconnect with comments like "This project is of no interest to me" and "It sucks".
- Concerns about the accuracy and completeness of answers provided by the tool: "He's not answering complete answers".
- In Sweden, students emphasized a desire for features supporting collaborative learning and peer interaction.
- Demand for a mobile version of the tool to make it more accessible.

3.3.3. Chat with Study Buddy

The analysis of student feedback regarding the chatting functionalities of SB across Greece, Cyprus, Sweden, and Ireland reveals a nuanced picture of its strengths and areas needing improvement.

Positive Aspects:

- Many students found the written and voice interaction features beneficial.

Co-funded by
the European Union

- The ease of using these features was generally praised, with comments such as "It's a nice app" and "Everything worked fine regardless if it wasn't useful".
- The Speech to Text and Text to Speech features were highly appreciated, especially for their benefits to children with vision problems.

Students' Suggestions for Improvement:

- More concise responses: "AI needs to describe some things more concisely".
- Faster responses: "Answer faster and don't stop for a long time".
- Enhancing the AI's understanding and responsiveness: "Doesn't answer the questions completely".
- More voice options, including both male and female voices, to enhance personalization.

Students' Critical Feedback:

- Some students found the tool struggled to understand what they were saying.
- Concerns about the artificial nature of the voice.
- Frustration with incomplete or irrelevant answers.

3.3.4. Chat History

The analysis of user feedback on the Chat History feature of SB from Greece, Cyprus, Sweden, and Ireland reveals several key insights into its performance and areas needing improvement.

Positive Aspects:

- In Cyprus, users found the Chat History feature very nice and useful with accurate responses.
- In Sweden, students appreciated the accuracy of the short titles summarizing their previous chats.
- In Ireland, feedback was overwhelmingly positive with students expressing satisfaction with the accuracy and ease of navigating the Chat History.

Students' Suggestions for Improvement:

- Improving the accuracy of chat summaries: "Could be more precise".
- Simplifying the navigation process and providing clearer guidance.
- Ensuring responses are not provided in English to enhance understanding and usability.

Students' Critical Feedback:

- In Greece, some users found the navigation moderately easy but noted a need for better interface design.

Co-funded by
the European Union

- Issues with the accuracy of information provided by SB, with some students rating it as only moderate or even inaccurate.

3.3.5. Usability, Functionality, and Impact on Learning

The overall feedback on the SB tool, drawn from survey responses across Greece, Cyprus, Sweden, and Ireland, indicates that it is generally well-received but there are significant areas that require attention and improvement.

Positive Aspects:

- Users appreciated several aspects of the tool, particularly its graphical user interface (GUI), ease of navigation, and overall usability.
- In Cyprus, respondents noted the effectiveness and potential of SB to improve students' performance.
- In Ireland, feedback indicated high praise for SB's performance, even surpassing ChatGPT according to some users.

Students' Suggestions for Improvement:

- Enhancing the GUI to make it more interactive and user-friendly.
- Improving the tool's functionality and providing clearer demonstrations of its benefits.
- Simplifying complex features and providing better guidance.
- Making the tool available on mobile phones, especially iPhones and other smartphones.
- Increasing the accuracy and completeness of answers provided by SB.

Students' Critical Feedback:

- Some users in Greece felt a disconnect, with comments like "I don't know," "nothing," and "I have no comment".
- Issues with the tool's responsiveness and comprehensiveness: "It doesn't answer all the questions".
- Suggestions for making responses faster and using more information from different sources.

3.3.6. Summary of findings from the qualitative research on Study Buddy – Correlations with quantitative analysis

The following table presents in a concise way the most prominent findings of the qualitative analysis of students' responses after testing and evaluating the SB through open-ended questions.

Table 31: SB Qualitative Research Results

Section	Existing Features	Weaknesses	Suggested Features

Co-funded by the European Union

Experience Using the SB Tools	Helpful tools, easy to use, well-designed, interactive learning, text summarization, correction tool, grading tool	Especially in non-English books, AI struggles with misspelled words, off-topic answers, discrepancies with book/internet info, inconsistent AI responses	More books, foreign language tools, self-control features, math problem solver, concise answers, broader subject coverage, intuitive UI, real-time language feedback, multimodal tool for images, avatar guide, immersive AR, "navigate" tool
Test and Student Dashboard	Easy to use, impactful on learning, efficient, multimodal features, presentations, sharing results, recording	Disconnect with students' needs, incomplete answers, UI issues, accuracy and relevance of questions	More diverse resources and questions, integrated communication features, collaborative learning tools, local educational content
Chat with SB	Beneficial written and voice interactions, intuitive design, user-friendly interface	Voice interactions not useful to some, speech recognition accuracy, inconsistent performance	Enhanced UI, improved voice recognition, comprehensive tutorials, faster responses, concise answers, additional voice options, interactive tools, better feedback mechanisms
Chat History	Useful feature, easy navigation, accurate short titles	Navigation difficulties, moderate/inaccurate information, inconsistent language in responses	More precise chat titles, simplified navigation, reliable and accurate responses, language consistency
Usability, Functionality, and Impact on Learning	Visually appealing GUI, ease of navigation, overall usability, positive interaction experiences	Complexity for some users, navigation challenges, inconsistent interactions, unclear daily task effectiveness	Improved GUI interactivity, clearer benefits, simplified tools, better onboarding, intuitive navigation aids, enhanced educational content, mobile version for iPhones and smartphones

Aligning these results with the results of the quantitative analysis provides some interesting **correlations** that shed light on the participants responses in closed-ended questions.

Regarding **user interface and ease of use**, the user-friendly and intuitive nature of the SB tools is consistently highlighted in both qualitative and quantitative feedback. Students from Cyprus and Ireland particularly praised the graphical user interface (GUI), with Cyprus showing 90% positive feedback and Ireland 100% in Q13a. This aligns with qualitative comments appreciating the visually appealing and functional design. However, Greece showed a significant portion of undecided users, suggesting that while the GUI meets basic expectations, there is room for improvement in making it more engaging and intuitive.

Tool complexity is noted in both sets of data. The qualitative feedback noted that some users find the tools unnecessarily complex, with Greece and Ireland expressing significant concerns in Q13c (56.14% and 64.71%, respectively). This suggests that simplifying tool functionalities

Co-funded by the European Union

and providing clearer instructions could enhance user experience and reduce perceived complexity.

Navigation within the SB tool received generally positive feedback, particularly from Cyprus and Ireland, both showing 100% positive responses in Q13d. However, the qualitative feedback indicated that some users still experience navigation challenges. Enhancing navigational features and providing better onboarding experiences could address these issues, ensuring a seamless experience for all users.

High **engagement and** positive **learning experiences** were reported, particularly in Ireland and Cyprus. The quantitative results from Q13i show that Cyprus (90%) and Ireland (100%) found the learning experience interesting, which aligns with the qualitative feedback that praised the interactive learning and text summarization features. However, Greece's lower engagement scores suggest a need for making learning experiences more interactive and compelling.

Regarding SB's performance and responsiveness, the qualitative feedback highlighted the need for faster and more accurate responses, particularly in handling misspelled words and providing on-topic answers. This aligns with the lower performance improvement ratings seen in Greece in Q13q (25.4%). Addressing these issues by improving AI accuracy and response times could enhance the overall user experience.

Technical resources and user knowledge required to use SB were generally sufficient, with high scores in Cyprus and Ireland (Q13m and Q13n). The qualitative feedback supports this, noting that most users have the necessary resources and knowledge, but there are suggestions for clearer guidance and additional training to support all users effectively.

Regarding future use intentions, there is strong intent to continue using SB in Cyprus and Ireland, as reflected in Q13p with 100% positive responses from both countries. This correlates with the qualitative feedback, which expressed satisfaction and a desire to keep using the tool. Conversely, Greece showed a significant portion of undecided users, indicating a need for ongoing engagement and demonstration of the tool's benefits to solidify commitment.

In terms of **voice and written Interactions**, qualitative feedback on the Chat with SB feature highlighted the benefits of written and voice interactions, though there were concerns about the accuracy and usability of voice interactions. This is mirrored in the quantitative results, where ease of use and usefulness of chat features received varied responses. Enhancing the accuracy of speech recognition and providing comprehensive tutorials could improve these aspects.

Regarding the **test taking functionality and Student Dashboard**, the qualitative feedback indicated a disconnect with some students' needs, incomplete answers, and UI issues. Quantitative results also showed mixed feelings about the effectiveness of the SB tools in daily tasks, with varying responses across countries. Incorporating more diverse resources

Co-funded by
the European Union

and questions, as suggested, could cater to a broader range of learning needs and improve overall satisfaction.

The **Chat History** feature received mixed reviews regarding navigation and accuracy. This aligns with quantitative results showing varied ease of use and accuracy ratings. Simplifying navigation and ensuring reliable information could enhance the user experience.

Overall, usability is a strong point, with significant portions of users finding SB easy to use. However, some users still experience difficulties, suggesting that targeted improvements based on qualitative feedback could further enhance usability.

Co-funded by
the European Union

4. Teacher Mate Pilot results

4.1. Quantitative Analysis

In the following subsections, the quantitative data collected by teachers through questionnaires are analysed. The subsections follow the teachers' questionnaire structure:

- **Questionnaire sections 2-8**: Experience using the TM tools (Lesson Plan Creation, Quiz Creation, Presentation Creation, Grading, Teacher Material Creation, Concept Exploration, Test Generation)
- **Questionnaire section 9**: Teacher Dashboard
- **Questionnaire section 10**: Chat with Teacher Mate
- **Questionnaire section 11**: Chat history and evaluation of quality of TM responses
- **Questionnaire section 13**: General usability and technology acceptance questions

Each section contained closed-ended questions aiming to collect information of the TM's effectiveness, usability, and areas for enhancement.

At the end of each section of the questionnaire, teachers were asked to share additional observations or recommendations about the TM tools and functionalities. The feedback provided in these open-ended questions will be presented in **Section 4.2, Qualitative Analysis**, which complements the quantitative analysis presented here.

4.1.1. Experience Using the Teacher Mate Tools

This section examines the evaluation of the available TM tools, focusing on teachers' experiences and perceptions. The TM Tools examined in this section are:

- Lesson Plan Creation
- Quiz Creation
- Presentation Creation
- Grading
- Teacher Material Creation
- Concept Exploration
- Test Generation

Each of the above tools was evaluated on the basis of three key rating criteria, reflected in corresponding questions of the questionnaire:

- The usefulness of the TM Tool for designing engaging and interactive lessons.
- The ease of using the TM Tool.
- The relevance of the parameters (course, topic, activities, class, subject, special educational needs etc) included in the TM Tool, which guide the creation of prompts for generating responses.

Co-funded by
the European Union

The following analysis provides a detailed breakdown of their responses, highlighting overall user satisfaction and areas for potential improvement. The insights gained will contribute to refining the Lesson Plan tool, ensuring it meets the practical needs of educators and supports their efforts to create effective and engaging educational experiences.

4.1.1.1. Lesson Plan Creation

Usefulness of Lesson Plan Suggestions (6a)

The survey responses as to the usefulness of the lesson plan suggestions generated by TM reveal positive feedback. Out of 32 respondents, the majority (23) rated the suggestions as either "Very" useful or "Extremely" useful. Specifically, 18 respondents rated it "Very" useful, and 5 rated it "Extremely" useful. This high level of satisfaction indicates that the lesson plan suggestions are highly valued for creating engaging and interactive lessons.

A smaller number of respondents found the suggestions to be only moderately useful, with 4 individuals selecting "Moderate." Meanwhile, 5 respondents rated it negatively, with 1 marking "Not at All" and 4 marking "Slightly." The distribution suggests that while most find the suggestions highly effective, a minority perceives room for improvement.

Ease of Using the Lesson Plan Component (6b)

Responses regarding the ease of using the Lesson Plan component show a significant lean towards positive feedback. Out of the 32 respondents, 27 found the process either "Very" easy or "Extremely" easy. Specifically, 13 respondents rated it as "Very" easy, and 14 rated it "Extremely" easy. This overwhelming positive feedback highlights the intuitive and user-friendly nature of the Lesson Plan component.

A modest number of respondents, 4, rated the ease of use as "Moderate," indicating that some users might encounter occasional difficulties but generally find the tool manageable. Only 1 respondent rated the ease of using the Lesson Plan component as "Slightly" easy, and none rated it as "Not at All" easy, further underscoring the overall positive reception of the tool's usability.

Relevance of Included Parameters (6c)

The third aspect evaluated was the relevance of the parameters included (such as course, topic, activities, number of modules) for generating lesson plans. The responses are largely positive. Out of the 32 respondents, 22 found the parameters either "Very" relevant or "Extremely" relevant. Specifically, 15 respondents rated the parameters as "Very" relevant, and 7 rated them as "Extremely" relevant. A small group, 8 respondents, rated the parameters' relevance as "Moderate," suggesting that while generally useful, there could be additional factors or improvements that might enhance their applicability. The least favorable ratings included 2 respondents who found the parameters only "Slightly" relevant and none who found them "Not at All" relevant.

Co-funded by
the European Union

In summary, the data suggests a strong positive reception of the lesson plan tools evaluated in the survey. Most respondents view the lesson plan suggestions as very or extremely useful, find using the Lesson Plan component very or extremely easy, and consider the parameters for lesson plan generation to be very or extremely relevant. Despite this overall positivity, a small subset of users expressed moderate satisfaction or dissatisfaction, indicating potential areas for improvement in the tools' functionality or the relevance of the suggested parameters (Figure 1).

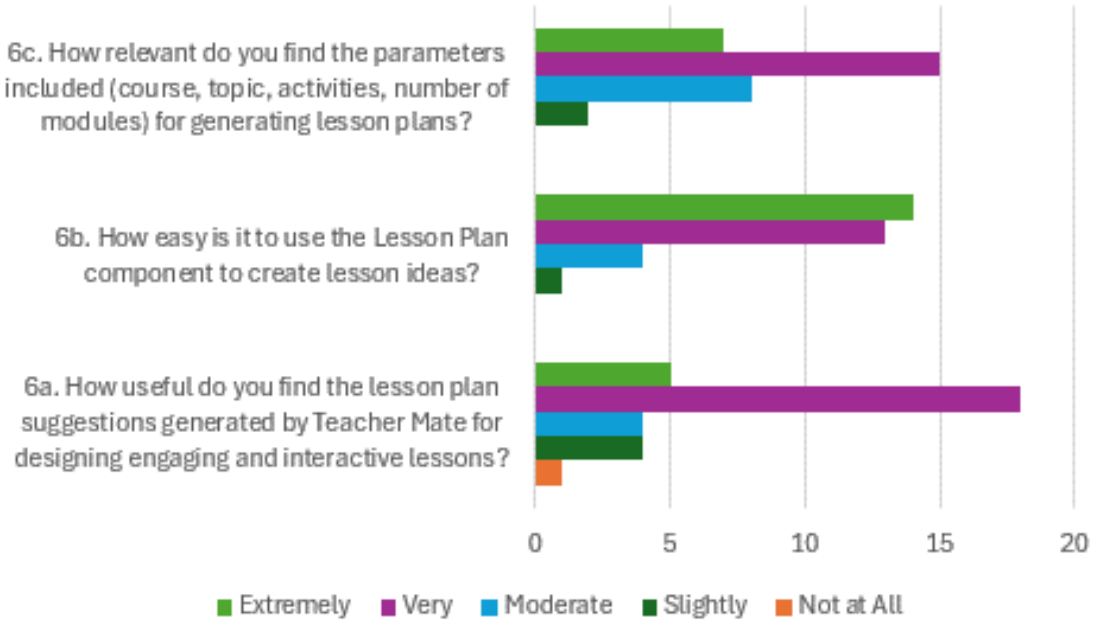

Figure 1: TM - Question 6 Results

4.1.1.2. Quiz Creation

Usefulness of the Create Quiz Component (9a)

The survey responses for the usefulness of the Create Quiz component for generating assessment quizzes reveal positive feedback. Out of 32 respondents, the majority (23) rated the component as either "Very" useful or "Extremely" useful. Specifically, 12 respondents rated it "Very" useful, and 11 rated it "Extremely" useful. This high level of satisfaction indicates that the Create Quiz component is highly valued for generating assessment quizzes.

A smaller number of respondents found the component to be only moderately useful, with 5 individuals selecting "Moderate." Meanwhile, 4 respondents rated it negatively, with 2 marking "Not at All" and another 2 marking "Slightly." The distribution suggests that while most find the component highly effective, a minority perceives room for improvement.

Co-funded by
the European Union

Ease of Creating Quizzes Using TM (9b)

Responses regarding the ease of creating quizzes using TM show a significant lean towards positive feedback. Out of the 32 respondents, 27 found the process either "Very" easy or "Extremely" easy. Specifically, 10 respondents rated it as "Very" easy, and 17 rated it "Extremely" easy. This overwhelming positive feedback highlights the intuitive and user-friendly nature of TM. A modest number of respondents, 4, rated the ease of use as "Moderate," indicating that some users might encounter occasional difficulties but generally find the tool manageable. Only 1 respondent rated the ease of creating quizzes as "Not at All" easy, and none rated it as "Slightly" easy, further underscoring the overall positive reception of the tool's usability.

Relevance of Suggested Parameters (9c)

The third aspect evaluated was the relevance of the suggested parameters (such as course, topics, number of matching questions) for generating quizzes. The responses are largely positive. Out of the 32 respondents, 25 found the parameters either "Very" relevant or "Extremely" relevant. Specifically, 13 respondents rated the parameters as "Very" relevant, and 12 rated them as "Extremely" relevant.

A small group, 6 respondents, rated the parameters' relevance as "Moderate," suggesting that while generally useful, there could be additional factors or improvements that might enhance their applicability. The least favorable ratings included 3 respondents who found the parameters only "Slightly" relevant and 1 who found them "Not at All" relevant. This indicates a minor portion of the user base may feel the suggested parameters do not fully align with their needs or expectations.

In summary, the data suggests a strong positive reception of the quiz creation tools evaluated in the survey. Most respondents view the Create Quiz component as very or extremely useful, find creating quizzes with TM very or extremely easy, and consider the suggested parameters for quiz generation to be very or extremely relevant. Despite this overall positivity, a small subset of users expressed moderate satisfaction or dissatisfaction, indicating potential areas for improvement in the tools' functionality or the relevance of the suggested parameters (Figure 2).

Co-funded by
the European Union

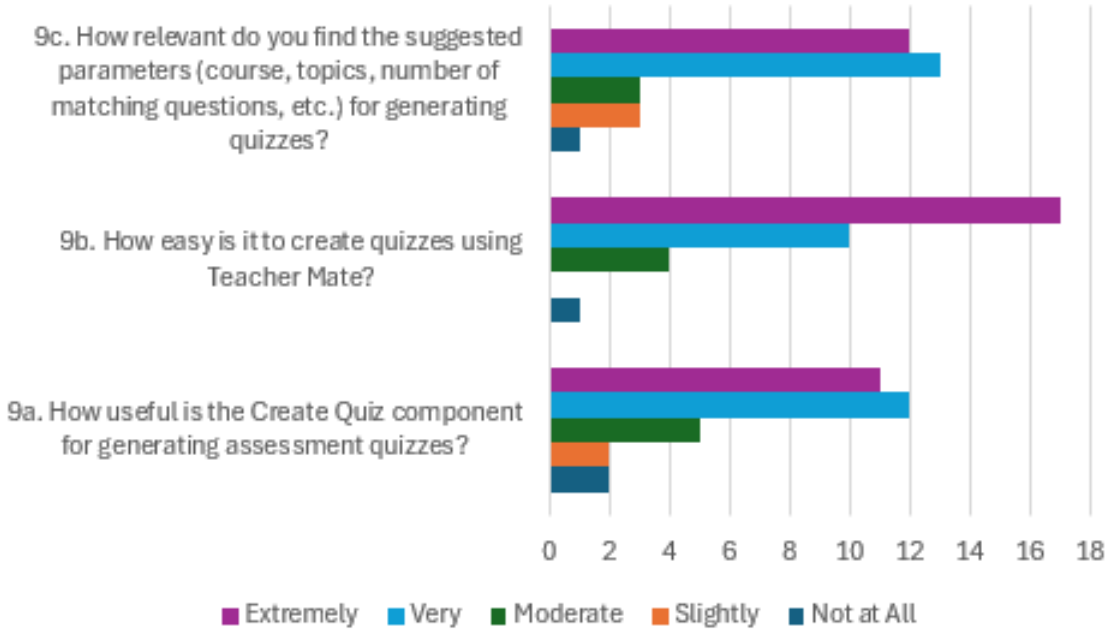

Figure 2: TM - Question 9 Results

4.1.1.3. Presentation Creation

Usefulness of the Presentation Design Ideas (12a)

The survey responses for the usefulness of the presentation design ideas generated by TM reveal positive feedback. Out of 32 respondents, the majority (22) rated the ideas as either "Very" useful or "Extremely" useful. Specifically, 17 respondents rated it "Very" useful, and 5 rated it "Extremely" useful. This high level of satisfaction indicates that the presentation design ideas are highly valued for creating engaging presentations.

A smaller number of respondents found the design ideas to be only moderately useful, with 6 individuals selecting "Moderate." Meanwhile, 4 respondents rated it negatively, with 1 marking "Not at All" and 3 marking "Slightly." The distribution suggests that while most find the design ideas highly effective, a minority perceives room for improvement.

Ease of Using the Create Presentation Component (12b)

Responses regarding the ease of using the Create Presentation component show a significant lean towards positive feedback. Out of the 32 respondents, 27 found the process either "Very" easy or "Extremely" easy. Specifically, 17 respondents rated it as "Very" easy, and 10 rated it "Extremely" easy. This overwhelming positive feedback highlights the intuitive and user-friendly nature of the Create Presentation component.

Co-funded by
the European Union

A modest number of respondents, 3, rated the ease of use as "Moderate," indicating that some users might encounter occasional difficulties but generally find the tool manageable. Only 1 respondent rated the ease of using the Create Presentation component as "Not at All" easy, and another 1 rated it as "Slightly" easy, further underscoring the overall positive reception of the tool's usability.

Relevance of Suggested Parameters (12c)

The third aspect evaluated was the relevance of the suggested parameters (such as course, topics, previous knowledge, duration) for generating presentation ideas. The responses are largely positive. Out of the 32 respondents, 24 found the parameters either "Very" relevant or "Extremely" relevant. Specifically, 19 respondents rated the parameters as "Very" relevant, and 5 rated them as "Extremely" relevant.

A small group, 8 respondents, rated the parameters' relevance as "Moderate," suggesting that while generally useful, there could be additional factors or improvements that might enhance their applicability. The least favorable ratings included 4 respondents who found the parameters only "Slightly" relevant and none who found them "Not at All" relevant. This indicates a minor portion of the user base may feel the suggested parameters do not fully align with their needs or expectations.

In summary, the data suggests a strong positive reception of the presentation design tools evaluated in the survey. Most respondents view the presentation design ideas as very or extremely useful, find using the Create Presentation component very or extremely easy, and consider the suggested parameters for presentation idea generation to be very or extremely relevant. Despite this overall positivity, a small subset of users expressed moderate satisfaction or dissatisfaction, indicating potential areas for improvement in the tools' functionality or the relevance of the suggested parameters (Figure 3).

Co-funded by
the European Union

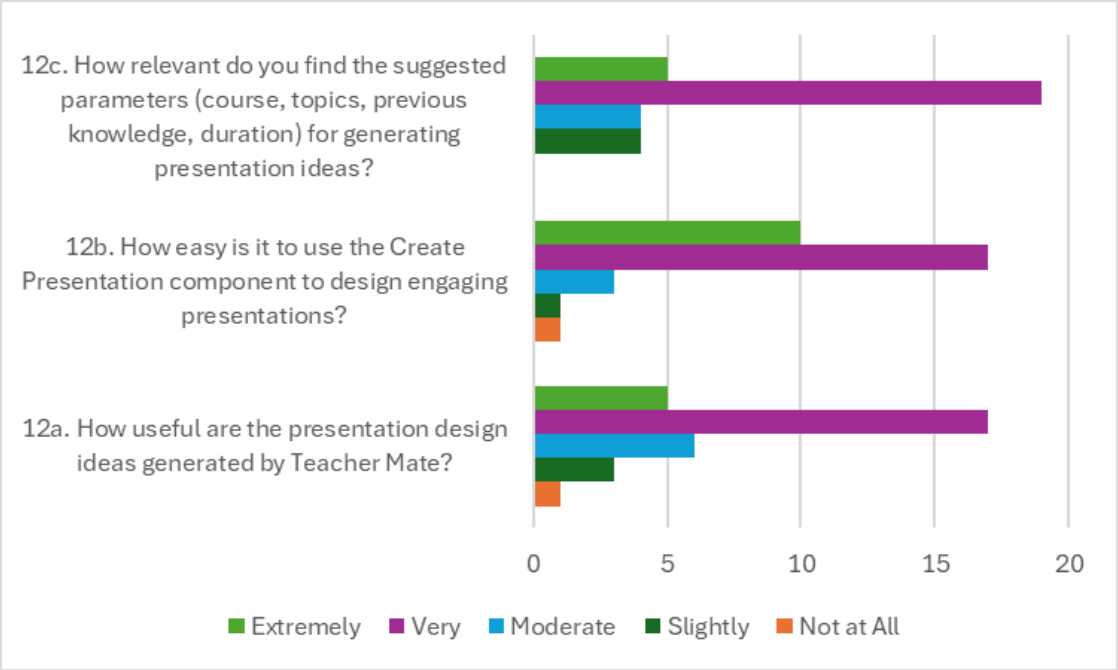

Figure 3: TM - Question 12 Results

4.1.1.4. Grading

Usefulness of the Grading Component (15a)

The survey responses for the usefulness of the Grading component for automatically grading student answers reveal positive feedback. Out of 32 respondents, the majority (21) rated the component as either "Very" useful or "Extremely" useful. Specifically, 11 respondents rated it "Very" useful, and 10 rated it "Extremely" useful. This high level of satisfaction indicates that the Grading component is highly valued for its effectiveness in automatically grading student answers.

A smaller number of respondents found the component to be only moderately useful, with 8 individuals selecting "Moderate." Meanwhile, 3 respondents rated it negatively, with 1 marking "Not at All" and 2 marking "Slightly." The distribution suggests that while most find the Grading component highly effective, a minority perceives room for improvement.

Ease of Using the Grading Component (15b)

Responses regarding the ease of using the Grading component show a significant lean towards positive feedback. Out of the 32 respondents, 24 found the process either "Very" easy or "Extremely" easy. Specifically, 13 respondents rated it as "Very" easy, and 11 rated it "Extremely" easy. This overwhelming positive feedback highlights the intuitive and user-friendly nature of the Grading component. A modest number of respondents, 6, rated the ease of use as "Moderate," indicating that some users might encounter occasional difficulties

Co-funded by
the European Union

but generally find the tool manageable. Only 1 respondent rated the ease of using the Grading component as "Not at All" easy, and another 1 rated it as "Slightly" easy, further underscoring the overall positive reception of the tool's usability.

Relevance of Suggested Parameters (15c)

The third aspect evaluated was the relevance of the suggested parameters (such as question, answer) for automated grading. The responses are largely positive. Out of the 32 respondents, 22 found the parameters either "Very" relevant or "Extremely" relevant. Specifically, 15 respondents rated the parameters as "Very" relevant, and 7 rated them as "Extremely" relevant.

A small group, 7 respondents, rated the parameters' relevance as "Moderate," suggesting that while generally useful, there could be additional factors or improvements that might enhance their applicability. The least favorable ratings included 3 respondents who found the parameters only "Slightly" relevant and 1 who found them "Not at All" relevant.

Accuracy of the AI Generated Mark (15d)

The survey responses regarding the accuracy of the AI-generated marks indicate a mix of positive and moderate feedback. Out of the 32 respondents, 18 found the AI-generated marks to be either "Very" accurate or "Extremely" accurate. Specifically, 11 respondents rated the marks as "Very" accurate, and 7 rated them as "Extremely" accurate. This suggests a good level of trust in the AI's ability to accurately grade student responses.

However, a notable number of respondents, 12, rated the accuracy as "Moderate," indicating that there are some reservations or inconsistencies perceived by users. Additionally, 8 respondents rated it negatively, with 2 marking "Not at All" accurate and 6 marking "Slightly" accurate, highlighting that accuracy is an area with potential for improvement.

Usefulness of TM's Explanatory Comments and Suggestions (15e)

The final aspect evaluated was the usefulness of TM's explanatory comments and suggestions for improvement. Out of the 32 respondents, a significant majority (23) found the comments and suggestions either "Very" useful or "Extremely" useful. Specifically, 15 respondents rated them as "Very" useful, and 8 rated them as "Extremely" useful. This indicates that the explanatory comments and suggestions are highly valued by users.

A smaller group, 7 respondents, rated the usefulness as "Moderate," suggesting that while generally appreciated, there may be areas for enhancing the explanatory feedback. Only 2 respondents rated the comments and suggestions negatively, with 1 marking "Not at All" useful and another 1 marking "Slightly" useful, indicating a minimal level of dissatisfaction.

In summary, the data suggests a strong positive reception of the grading tools and features evaluated in the survey. Despite this overall positivity, a small subset of users expressed

Co-funded by
the European Union

moderate satisfaction or dissatisfaction, indicating potential areas for improvement in the tools' functionality, accuracy of the AI-generated marks, or the relevance of the suggested parameters (Figure 4).

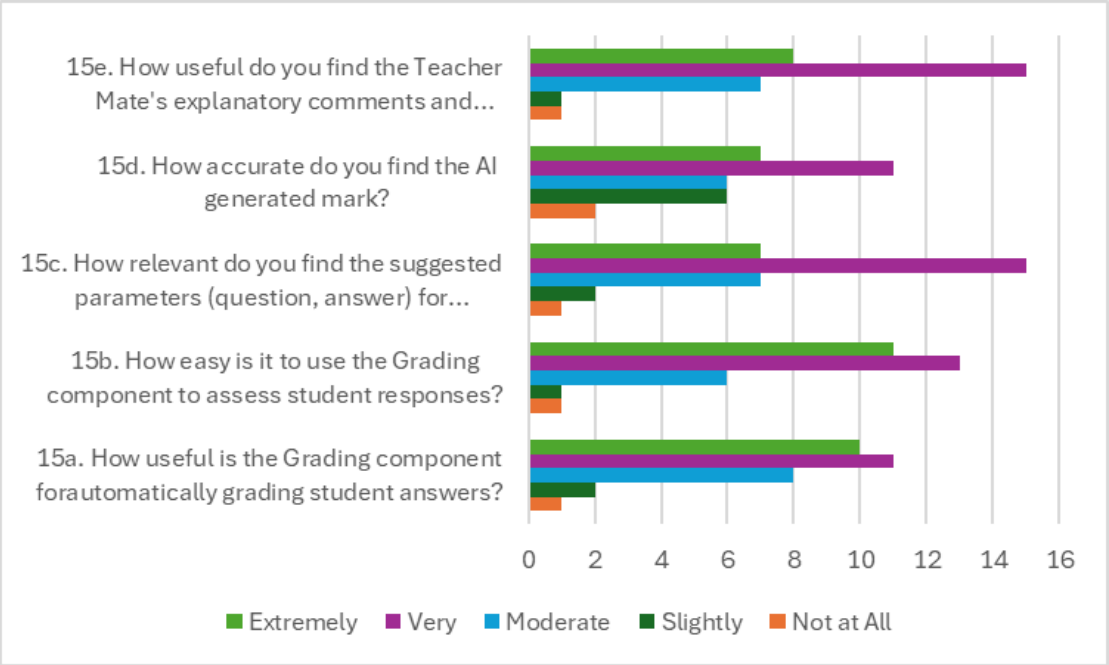

Figure 4: TM - Question 15 Results

4.1.1.5. Creating Teaching Material

Usefulness of the Component for Creating Supporting Material (18a)

The survey responses for the usefulness of the component for creating material that supports the study of a topic reveal positive feedback. Out of 32 respondents, the majority (22) rated the component as either "Very" useful or "Extremely" useful. Specifically, 16 respondents rated it "Very" useful, and 6 rated it "Extremely" useful. This high level of satisfaction indicates that the component is highly valued for its effectiveness in creating supporting study material.

A smaller number of respondents found the component to be only moderately useful, with 7 individuals selecting "Moderate." Meanwhile, 3 respondents rated it negatively, with 2 marking "Not at All" and 1 marking "Slightly." The distribution suggests that while most find the component highly effective, a minority perceives room for improvement.

Ease of Using the Create Teaching Material Component (18b)

Responses regarding the ease of using the Create Teaching Material component show a significant lean towards positive feedback. Out of the 32 respondents, 27 found the process either "Very" easy or "Extremely" easy. Specifically, 19 respondents rated it as "Very" easy,

Co-funded by
the European Union

and 8 rated it "Extremely" easy. This overwhelming positive feedback highlights the intuitive and user-friendly nature of the Create Teaching Material component.

A modest number of respondents, 2, rated the ease of use as "Moderate," indicating that some users might encounter occasional difficulties but generally find the tool manageable. Only 3 respondents rated the ease of using the Create Teaching Material component negatively, with 1 marking "Not at All" easy and 2 marking "Slightly" easy, further underscoring the overall positive reception of the tool's usability.

Relevance of Suggested Parameters (18c)

The third aspect evaluated was the relevance of the suggested parameters (such as topic, previous knowledge, special student needs) for explaining key points. The responses are largely positive. Out of the 32 respondents, 21 found the parameters either "Very" relevant or "Extremely" relevant. Specifically, 16 respondents rated the parameters as "Very" relevant, and 5 rated them as "Extremely" relevant.

A small group, 7 respondents, rated the parameters' relevance as "Moderate," suggesting that while generally useful, there could be additional factors or improvements that might enhance their applicability. The least favorable ratings included 4 respondents who found the parameters only "Slightly" relevant and none who found them "Not at All" relevant. This indicates a minor portion of the user base may feel the suggested parameters do not fully align with their needs or expectations.

In summary, the data suggests a strong positive reception of the teaching material creation tools evaluated in the survey. Despite this overall positivity, a small subset of users expressed moderate satisfaction or dissatisfaction, indicating potential areas for improvement in the tools' functionality or the relevance of the suggested parameters (Figure 5).

Co-funded by
the European Union

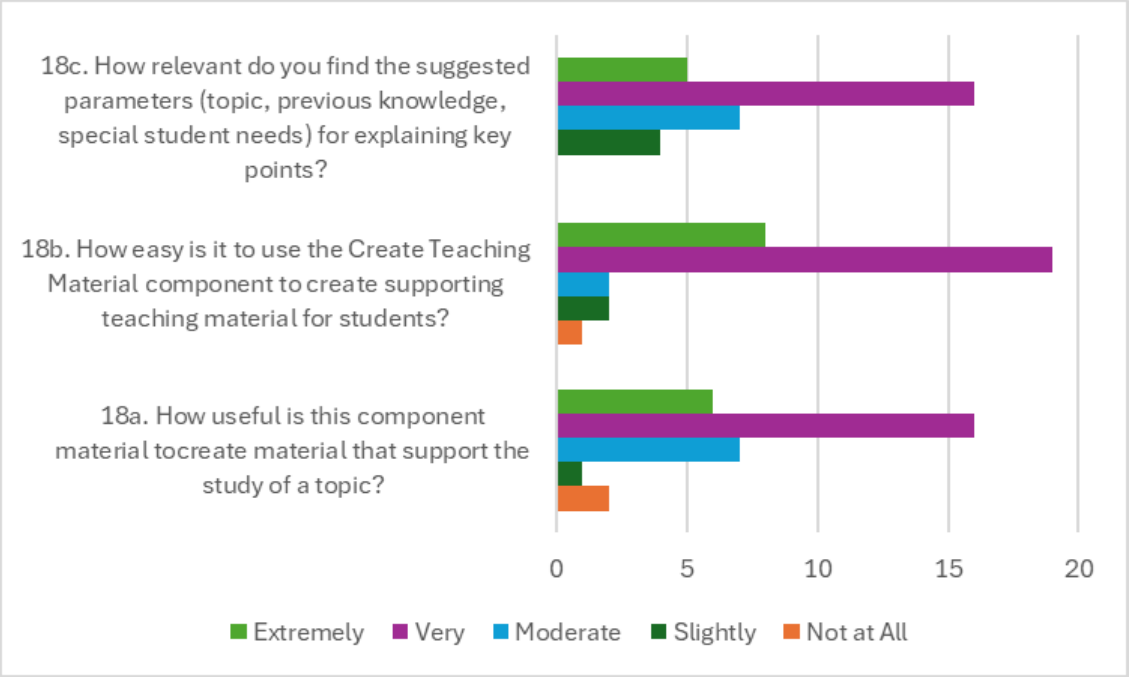

Figure 5: TM - Question 18 Results

4.1.1.6. Concept Exploration

Usefulness of the Concept Exploration Component (21a)

The survey responses for the usefulness of the Concept Exploration component for devising engaging ways to explain new concepts reveal positive feedback. Out of 32 respondents, the majority (22) rated the component as either "Very" useful or "Extremely" useful. Specifically, 18 respondents rated it "Very" useful, and 4 rated it "Extremely" useful. This high level of satisfaction indicates that the Concept Exploration component is highly valued for its effectiveness in explaining new concepts engagingly. A smaller number of respondents found the component to be only moderately useful, with 8 individuals selecting "Moderate." Meanwhile, 2 respondents rated it negatively, with 1 marking "Not at All" and 1 marking "Slightly." The distribution suggests that while most find the Concept Exploration component highly effective, a minority perceives room for improvement.

Ease of Using the Concept Exploration Component (21b)

Responses regarding the ease of using the Concept Exploration component show a significant lean towards positive feedback. Out of the 32 respondents, 26 found the process either "Very" easy or "Extremely" easy. Specifically, 18 respondents rated it as "Very" easy, and 8 rated it "Extremely" easy. This overwhelming positive feedback highlights the intuitive and user-friendly nature of the Concept Exploration component. A modest number of respondents, 5, rated the ease of use as "Moderate," indicating that some users might encounter occasional difficulties but generally find the tool manageable. Only 1 respondent

Co-funded by
the European Union

rated the ease of using the Concept Exploration component as "Not at All" easy, and none rated it as "Slightly" easy, further underscoring the overall positive reception of the tool's usability.

Relevance of Suggested Parameters (21c)

The third aspect evaluated was the relevance of the suggested parameters (such as terms, grade, special student needs) for concept exploration. The responses are largely positive. Out of the 32 respondents, 20 found the parameters either "Very" relevant or "Extremely" relevant. Specifically, 15 respondents rated the parameters as "Very" relevant, and 5 rated them as "Extremely" relevant.

A small group, 9 respondents, rated the parameters' relevance as "Moderate," suggesting that while generally useful, there could be additional factors or improvements that might enhance their applicability. The least favorable ratings included 3 respondents who found the parameters only "Slightly" relevant and none who found them "Not at All" relevant. This indicates a minor portion of the user base may feel the suggested parameters do not fully align with their needs or expectations.

In summary, the data suggests a strong positive reception of the concept exploration tools evaluated in the survey. Most respondents view the Concept Exploration component as very or extremely useful, find using the Concept Exploration component very or extremely easy, and consider the suggested parameters for concept exploration to be very or extremely relevant. Despite this overall positivity, a small subset of users expressed moderate satisfaction or dissatisfaction, indicating potential areas for improvement in the tools' functionality or the relevance of the suggested parameters (Figure 6).

Co-funded by
the European Union

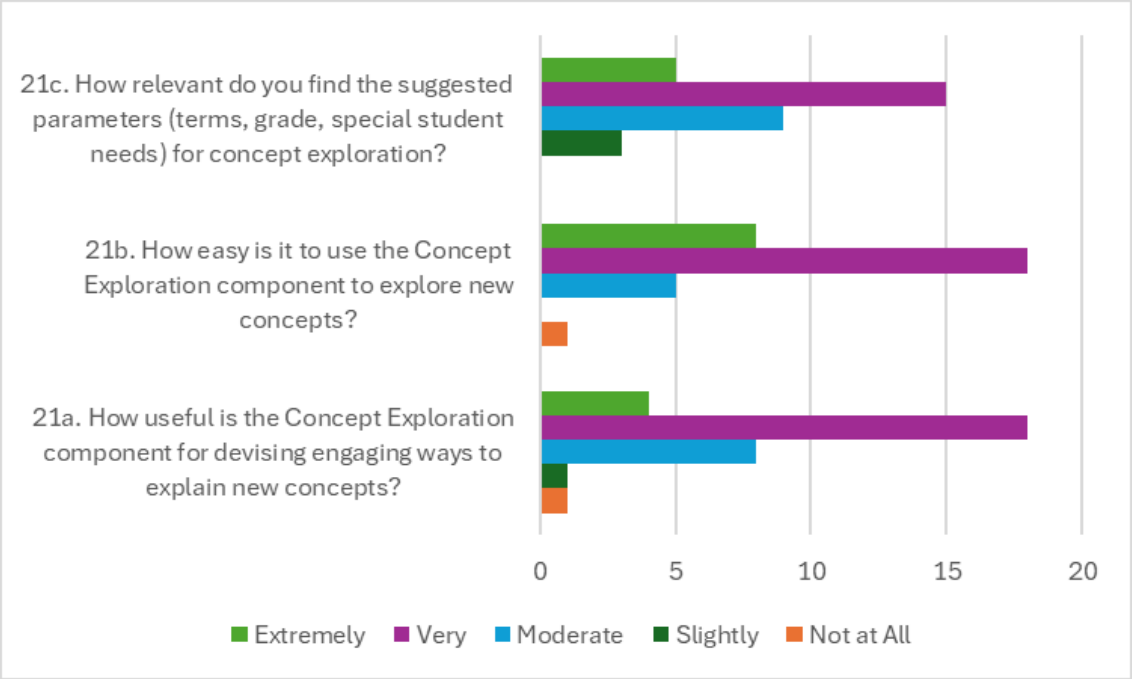

Figure 6: TM - Question 21 Results

4.1.1.7. Test Generation

Usefulness of the Generate Test Component (24a)

The survey responses for the usefulness of the Generate Test component for creating tests reveal positive feedback. Out of 29 respondents, the majority (22) rated the component as either "Very" useful or "Extremely" useful. Specifically, 15 respondents rated it "Very" useful, and 7 rated it "Extremely" useful. This high level of satisfaction indicates that the Generate Test component is highly valued for its effectiveness in creating tests. A smaller number of respondents found the component to be only moderately useful, with 5 individuals selecting "Moderate." Meanwhile, 2 respondents rated it negatively, with 1 marking "Not at All" and 1 marking "Slightly." The distribution suggests that while most find the Generate Test component highly effective, a minority perceives room for improvement.

Ease of Using the Generate Test Component (24b)

Responses regarding the ease of using the Generate Test component show a significant lean towards positive feedback. Out of the 29 respondents, 23 found the process either "Very" easy or "Extremely" easy. Specifically, 13 respondents rated it as "Very" easy, and 10 rated it "Extremely" easy. This overwhelming positive feedback highlights the intuitive and user-friendly nature of the Generate Test component.

A modest number of respondents, 5, rated the ease of use as "Moderate," indicating that some users might encounter occasional difficulties but generally find the tool manageable.

Co-funded by
the European Union

Only 1 respondent rated the ease of using the Generate Test component as "Slightly" easy, and none rated it as "Not at All" easy, further underscoring the overall positive reception of the tool's usability.

Relevance of Suggested Parameters (24c)

The third aspect evaluated was the relevance of the suggested parameters (such as test title, description, total marks, duration, question types, due date) for generating tests. The responses are largely positive. Out of the 29 respondents, 22 found the parameters either "Very" relevant or "Extremely" relevant. Specifically, 14 respondents rated the parameters as "Very" relevant, and 8 rated them as "Extremely" relevant. A small group, 6 respondents, rated the parameters' relevance as "Moderate," suggesting that while generally useful, there could be additional factors or improvements that might enhance their applicability. The least favorable ratings included 1 respondent who found the parameters not at all relevant and none who found them only "Slightly" relevant. This indicates a minor portion of the user base may feel the suggested parameters do not fully align with their needs or expectations.

Accuracy of the Information Provided (24d)

The survey responses regarding the accuracy of the information provided in the generated questions and correct answers of the test indicate a mix of positive and moderate feedback. Out of the 29 respondents, 18 found the information to be either "Very" accurate or "Extremely" accurate. Specifically, 12 respondents rated the accuracy as "Very" accurate, and 6 rated it as "Extremely" accurate. This suggests a good level of trust in the tool's ability to provide accurate information. However, a notable number of respondents, 8, rated the accuracy as "Moderate," indicating that there are some reservations or inconsistencies perceived by users. Additionally, 3 respondents rated it negatively, with 2 marking "Not at All" accurate and 1 marking "Slightly" accurate, highlighting that accuracy is an area with potential for improvement.

Correctness of the Language Used (24e)

The survey responses regarding the correctness of the language used in the generated questions and correct answers indicate a mix of positive and moderate feedback. Out of the 29 respondents, 18 found the language to be either "Very" correct or "Extremely" correct. Specifically, 10 respondents rated the language as "Very" correct, and 8 rated it as "Extremely" correct. This suggests a good level of satisfaction with the language used. However, a notable number of respondents, 7, rated the correctness of the language as "Moderate," indicating that there are some areas for improvement. Additionally, 4 respondents rated it negatively, with 2 marking "Not at All" correct and 2 marking "Slightly" correct, suggesting some dissatisfaction with the language used in the generated content.

Co-funded by
the European Union

Relevance to the Test Subject and Description (24f)

The survey responses regarding the relevance of the generated questions and correct answers to the test subject and description are largely positive. Out of the 29 respondents, 20 found the relevance either "Very" relevant or "Extremely" relevant. Specifically, 14 respondents rated the relevance as "Very" relevant, and 6 rated it as "Extremely" relevant. A smaller group, 6 respondents, rated the relevance as "Moderate," suggesting that while generally useful, there could be additional factors or improvements that might enhance their applicability. The least favorable ratings included 3 respondents who found the relevance not at all or only slightly relevant, indicating a minor portion of the user base may feel the generated content does not fully align with their expectations.

Usefulness of the Editing Option (24g)

The survey responses regarding the usefulness of the editing option for generated questions, correct answers, and marks per question are highly positive. Out of the 29 respondents, 23 found the editing option either "Very" useful or "Extremely" useful. Specifically, 11 respondents rated the editing option as "Very" useful, and 12 rated it as "Extremely" useful. A smaller group, 3 respondents, rated the usefulness of the editing option as "Moderate," suggesting that while generally appreciated, there could be some areas for improvement. Only 2 respondents rated it negatively, with 1 marking "Not at All" useful and 1 marking "Slightly" useful, indicating minimal dissatisfaction.

Ease of Editing Generated Questions and Answers (24h)

The survey responses regarding the ease of editing generated questions and correct answers are highly positive. Out of the 29 respondents, 24 found the process either "Very" easy or "Extremely" easy. Specifically, 14 respondents rated it as "Very" easy, and 10 rated it "Extremely" easy. This overwhelming positive feedback highlights the intuitive and user-friendly nature of the editing process. A smaller group, 4 respondents, rated the ease of editing as "Moderate," indicating that while generally easy, some users might encounter occasional difficulties. Only 1 respondent rated the ease of editing negatively, marking it as "Slightly" easy, and none rated it as "Not at All" easy, further underscoring the overall positive reception of the tool's usability.

In summary, the data suggests a strong positive reception of the test generation and editing tools evaluated in the survey. Most respondents view the Generate Test component as very or extremely useful, find using and editing the component very or extremely easy, and consider the suggested parameters for test generation to be very or extremely relevant. Additionally, most respondents find the information provided, the language used, and the relevance to the test subject to be accurate and appropriate. Despite this overall positivity, a small subset of users expressed moderate satisfaction or dissatisfaction, indicating potential areas for improvement in the tools' functionality, accuracy of the generated content, and the relevance of the suggested parameters (Figure 7).

Co-funded by
the European Union

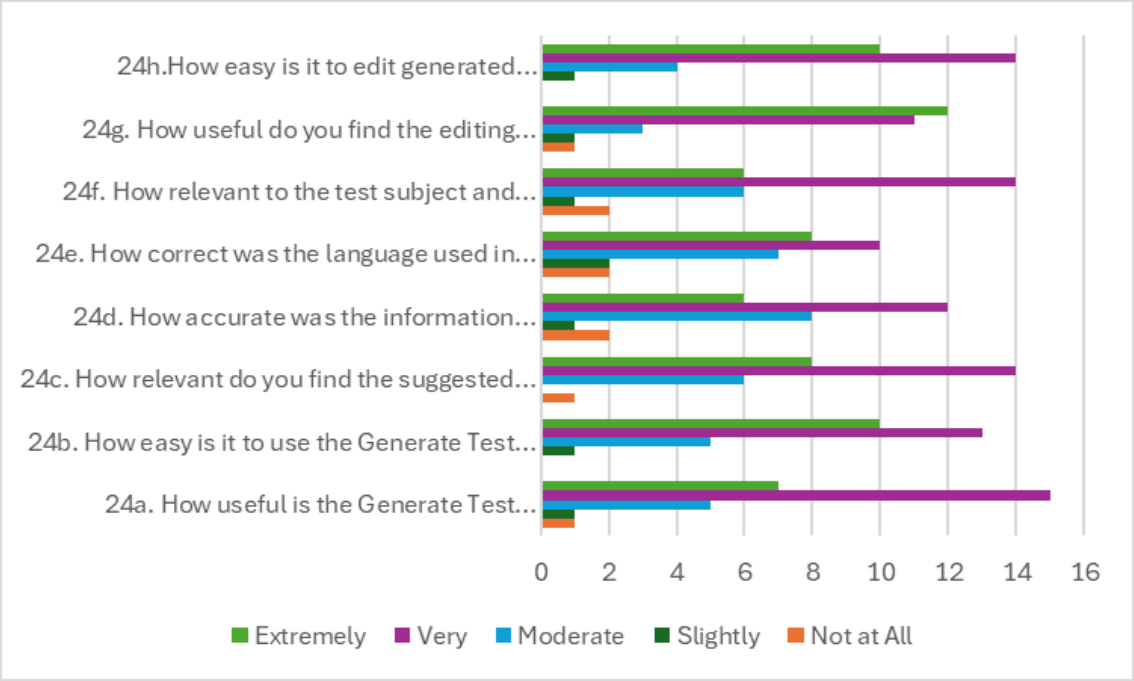

Figure 7: TM - Question 24 Results

4.1.2. Experience of the Teacher Dashboard functionality

This section presents the evaluation of the Teacher Dashboard component of the TM conversational AI assistant, focusing on teachers' experiences and feedback. The questionnaire section related to the Teacher Dashboard included ten key rating questions, to evaluate:

- The usefulness of the Add class component for creating a class.
- The ease of using the Add class component to create a class.
- The relevance of the suggested parameters (course name, grade, number of students) for creating a class.
- The helpfulness of the visualizations (diagrams) about overall class performance (total grades, last test grades, time spent on last test).
- The helpfulness of the visualizations (diagrams) and the rest of the information about individual student performance and progress (average grade, last test grade, % of tests completed).
- The usefulness of the Update marks functionality.
- The ease of reviewing individual students' answers to the test.
- The accuracy of the AI-generated marks.
- The usefulness of the AI-generated Student Overall Assessment.
- The usefulness of the AI-generated Test Feedback.

The subsequent analysis will provide a detailed examination of the responses to these questions, highlighting overall user satisfaction and identifying areas for potential

Co-funded by
the European Union

improvement. The insights obtained will be exploited in refining the Teacher Dashboard component, ensuring it meets educators' needs and supports their efforts in effectively managing and accessing class and student performance.

Usefulness of the Add Class Component (27a)

The survey responses for the usefulness of the Add Class component for creating a class reveal positive feedback. Out of 32 respondents, the majority (28) rated the component as either "Very" useful or "Extremely" useful. Specifically, 13 respondents rated it "Very" useful, and 15 rated it "Extremely" useful. This high level of satisfaction indicates that the Add Class component is highly valued for its effectiveness in creating classes. A smaller number of respondents found the component to be only moderately useful, with 2 individuals selecting "Moderate." There were no respondents who rated it negatively, with none marking "Not at All" or "Slightly." The distribution suggests that most find the Add Class component highly effective.

Ease of Using the Add Class Component (27b)

Responses regarding the ease of using the Add Class component show a significant lean towards positive feedback. Out of the 32 respondents, 31 found the process either "Very" easy or "Extremely" easy. Specifically, 16 respondents rated it as "Very" easy, and 15 rated it "Extremely" easy. This overwhelming positive feedback highlights the intuitive and user-friendly nature of the Add Class component. Only 1 respondent rated the ease of use as "Moderate," indicating that some users might encounter occasional difficulties but generally find the tool manageable. There were no negative ratings, further underscoring the overall positive reception of the tool's usability.

Relevance of Suggested Parameters (27c)

The third aspect evaluated was the relevance of the suggested parameters (such as course name, grade, number of students) for creating a class. The responses are largely positive. Out of the 32 respondents, 30 found the parameters either "Very" relevant or "Extremely" relevant. Specifically, 18 respondents rated the parameters as "Very" relevant, and 12 rated them as "Extremely" relevant. A small group, 2 respondents, rated the parameters' relevance as "Moderate," suggesting that while generally useful, there could be additional factors or improvements that might enhance their applicability. There were no respondents who found the parameters not at all or only slightly relevant, indicating strong alignment with user needs.

Helpfulness of Class Performance Visualizations (27d)

The survey responses regarding the helpfulness of visualizations (diagrams) about overall class performance (Total grades, Last test grades, Time spent on last test) are highly positive. Out of the 32 respondents, 28 found the visualizations either "Very" helpful or "Extremely" helpful. Specifically, 14 respondents rated the visualizations as "Very" helpful, and 14 rated them as "Extremely" helpful. A smaller group, 3 respondents, rated the visualizations as

Co-funded by
the European Union

"Moderate," suggesting that while generally helpful, there could be areas for improvement. Only 1 respondent rated the visualizations slightly helpful, and none rated them not at all helpful, indicating overall satisfaction with the provided information.

Helpfulness of Individual Student Performance Visualizations (27e)

The survey responses regarding the helpfulness of visualizations and information about individual student performance and progress (average grade, last test grade, % of tests completed) are highly positive. Out of the 32 respondents, 28 found the visualizations either "Very" helpful or "Extremely" helpful. Specifically, 16 respondents rated the visualizations as "Very" helpful, and 12 rated them as "Extremely" helpful. A smaller group, 4 respondents, rated the visualizations as "Moderate," suggesting that while generally helpful, there could be areas for improvement. Only 1 respondent rated the visualizations slightly helpful, and none rated them not at all helpful, indicating overall satisfaction with the provided information.

Usefulness of the Update Marks Functionality (27f)

The survey responses regarding the usefulness of the Update Marks functionality are highly positive. Out of the 32 respondents, 29 found the functionality either "Very" useful or "Extremely" useful. Specifically, 13 respondents rated it as "Very" useful, and 16 rated it as "Extremely" useful. A smaller group, 2 respondents, rated the functionality as "Moderate," suggesting that while generally useful, there could be areas for improvement. Only 1 respondent rated it slightly useful, and none rated it not at all useful, indicating minimal dissatisfaction.

Ease of Reviewing Individual Students' Answers (27g)

The survey responses regarding the ease of reviewing individual students' answers to the test are highly positive. Out of the 32 respondents, 29 found the process either "Very" easy or "Extremely" easy. Specifically, 17 respondents rated it as "Very" easy, and 12 rated it "Extremely" easy. A smaller group, 3 respondents, rated the ease of reviewing as "Moderate," indicating that while generally easy, some users might encounter occasional difficulties. There were no negative ratings, further underscoring the overall positive reception of the tool's usability.

Accuracy of the AI Generated Marks (27h)

The survey responses regarding the accuracy of the AI-generated marks indicate a mix of positive and moderate feedback. Out of the 32 respondents, 24 found the AI-generated marks to be either "Very" accurate or "Extremely" accurate. Specifically, 18 respondents rated the marks as "Very" accurate, and 6 rated them as "Extremely" accurate. However, a notable number of respondents, 8, rated the accuracy as "Moderate," indicating that there are some reservations or inconsistencies perceived by users. Additionally, 2 respondents rated it

Co-funded by
the European Union

negatively, marking "Slightly" accurate, highlighting that accuracy is an area with potential for improvement.

Usefulness of the AI-Generated Student Overall Assessment (27i)

The survey responses regarding the usefulness of the AI-generated Student Overall Assessment are highly positive. Out of the 32 respondents, 27 found the assessment either "Very" useful or "Extremely" useful. Specifically, 18 respondents rated the assessment as "Very" useful, and 9 rated it as "Extremely" useful. A smaller group, 3 respondents, rated the assessment as "Moderate," suggesting that while generally useful, there could be areas for improvement. Only 2 respondents rated it slightly useful, and none rated it not at all useful, indicating overall satisfaction with the AI-generated assessment.

Usefulness of the AI-Generated Test Feedback (27j)

The survey responses regarding the usefulness of the AI-generated Test Feedback are highly positive. Out of the 32 respondents, 28 found the feedback either "Very" useful or "Extremely" useful. Specifically, 20 respondents rated the feedback as "Very" useful, and 8 rated it as "Extremely" useful. A smaller group, 1 respondent, rated the feedback as "Moderate," suggesting that while generally useful, there could be areas for improvement. Only 3 respondents rated it slightly useful, and none rated it not at all useful, indicating overall satisfaction with the AI-generated feedback.

In summary, the data suggests a strong positive reception of the various components and functionalities of the Teacher Dashboard. Most respondents view the Add Class component, visualizations of class and individual performance, Update Marks functionality, and AI-generated assessments and feedback as very or extremely useful. They also find these components very or extremely easy to use and relevant. Despite this overall positivity, a small subset of users expressed moderate satisfaction or dissatisfaction, indicating potential areas for improvement in the tools' functionality, accuracy of the AI-generated marks, and relevance of the suggested parameters (Figure 8).

Co-funded by
the European Union

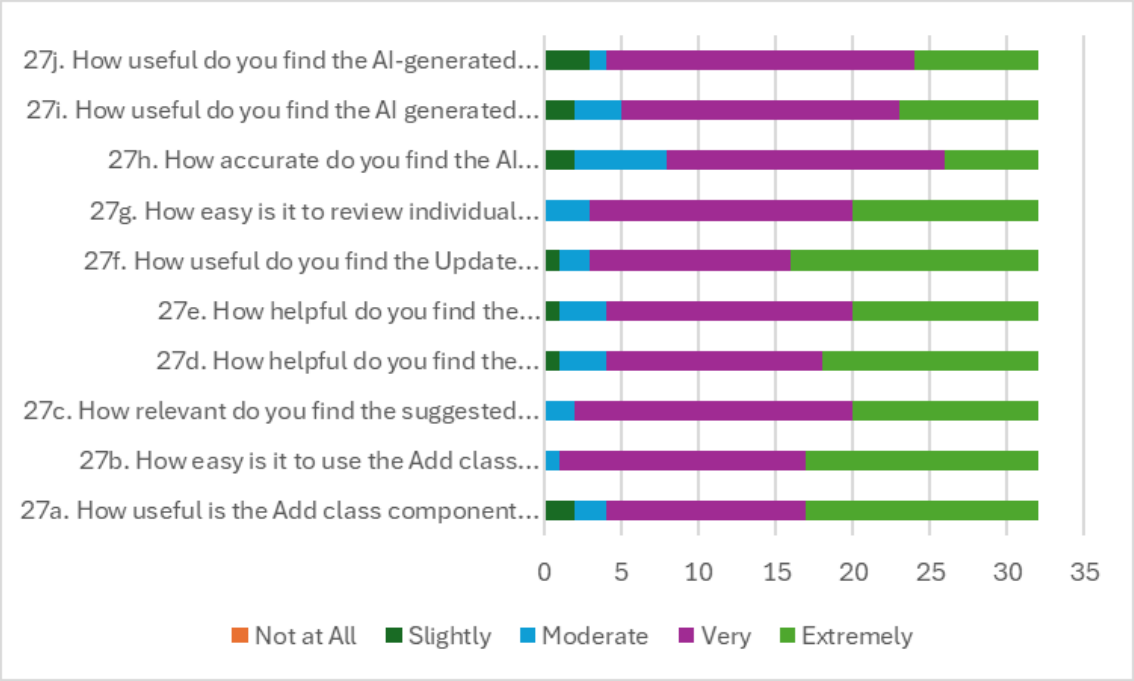

Figure 8: TM - Question 27 Results

4.1.3. Experience on Chatting with the Teacher Mate

This section presents the evaluation of the Chat with TM functionality of the TM conversational AI assistant, focusing on teachers' experiences and feedback. The questionnaire included six key rating questions of the chatting functionality:

- The usefulness of chatting with TM in writing.
- The ease of chatting with TM in writing.
- The usefulness of chatting with TM using voice interaction.
- The ease of chatting with TM using voice interaction.
- The quality of speech recognition.
- The quality of speech generation.

The subsequent analysis will provide a detailed examination of the responses to these questions, highlighting overall user satisfaction and identifying areas for potential improvement regarding both text and voice communication.

Usefulness of Chatting with TM in Writing (31a)

The survey responses for the usefulness of chatting with TM in writing reveal positive feedback. Out of 32 respondents, the majority (25) rated the experience as either "Very" useful or "Extremely" useful. Specifically, 15 respondents rated it "Very" useful, and 10 rated it "Extremely" useful. This high level of satisfaction indicates that chatting with TM in writing is highly valued. A smaller number of respondents found the experience to be only moderately

Co-funded by
the European Union

useful, with 5 individuals selecting "Moderate." Only 2 respondents rated it negatively, with 2 marking "Slightly." None of the respondents rated it "Not at All" useful, suggesting a generally favorable view of this feature.

Ease of Chatting with TM in Writing (31b)

Responses regarding the ease of chatting with TM in writing show a significant lean towards positive feedback. Out of the 32 respondents, 27 found the process either "Very" easy or "Extremely" easy. Specifically, 17 respondents rated it as "Very" easy, and 10 rated it "Extremely" easy. This overwhelming positive feedback highlights the intuitive and user-friendly nature of this feature. A smaller number of respondents, 3, rated the ease of use as "Moderate," indicating that some users might encounter occasional difficulties but generally find the tool manageable. Only 2 respondents rated it negatively, with 1 marking "Not at All" easy and 1 marking "Slightly" easy, further underscoring the overall positive reception of the tool's usability.

Usefulness of Chatting with TM Using Voice Interaction (31c)

The survey responses for the usefulness of chatting with TM using voice interaction are highly positive. Out of 32 respondents, the majority (22) rated the experience as either "Very" useful or "Extremely" useful. Specifically, 7 respondents rated it "Very" useful, and 15 rated it "Extremely" useful. This indicates that voice interaction with TM is highly valued. A smaller number of respondents found the experience to be only moderately useful, with 7 individuals selecting "Moderate." Only 3 respondents rated it negatively, with 1 marking "Not at All" and 2 marking "Slightly." The distribution suggests that while most find the voice interaction highly effective, there is a minority who perceive room for improvement.

Ease of Chatting with TM Using Voice Interaction (31d)

Responses regarding the ease of chatting with TM using voice interaction show a significant lean towards positive feedback. Out of the 32 respondents, 25 found the process either "Very" easy or "Extremely" easy. Specifically, 8 respondents rated it as "Very" easy, and 17 rated it "Extremely" easy. This positive feedback highlights the user-friendly nature of voice interaction. A modest number of respondents, 5, rated the ease of use as "Moderate," indicating occasional difficulties but general manageability. Only 2 respondents rated it negatively, with 1 marking "Not at All" easy and 1 marking "Slightly" easy, further emphasizing the overall positive reception of the tool's usability.

Quality of Speech Recognition (31e)

The survey responses regarding the quality of speech recognition indicate positive feedback. Out of 32 respondents, the majority (26) rated the quality as either "Very" good or "Extremely" good. Specifically, 14 respondents rated it "Very" good, and 12 rated it "Extremely" good. This high level of satisfaction suggests that the speech recognition feature performs well. A smaller number of respondents found the quality to be only moderate, with 5 individuals

Co-funded by
the European Union

selecting "Moderate." Only 1 respondent rated the quality slightly good, and none rated it not at all good, indicating overall satisfaction with the speech recognition quality.

Quality of Speech Generation (31f)

The survey responses regarding the quality of speech generation are largely positive. Out of the 32 respondents, the majority (24) rated the quality as either "Very" good or "Extremely" good. Specifically, 9 respondents rated it "Very" good, and 15 rated it "Extremely" good. This indicates a high level of satisfaction with the speech generation feature. A smaller number of respondents found the quality to be only moderate, with 7 individuals selecting "Moderate." Only 1 respondent rated the quality slightly good, and none rated it not at all good, indicating overall positive feedback.

In summary, the data suggests a strong positive reception of the various chat and voice interaction features of TM. Most respondents view chatting with TM in writing and using voice interaction as very or extremely useful and easy. They also find the quality of speech recognition and generation to be high, and the short titles of previous chats to be accurate. Despite this overall positivity, a small subset of users expressed moderate satisfaction or dissatisfaction, indicating potential areas for improvement in the tools' functionality and quality (Figure 9).

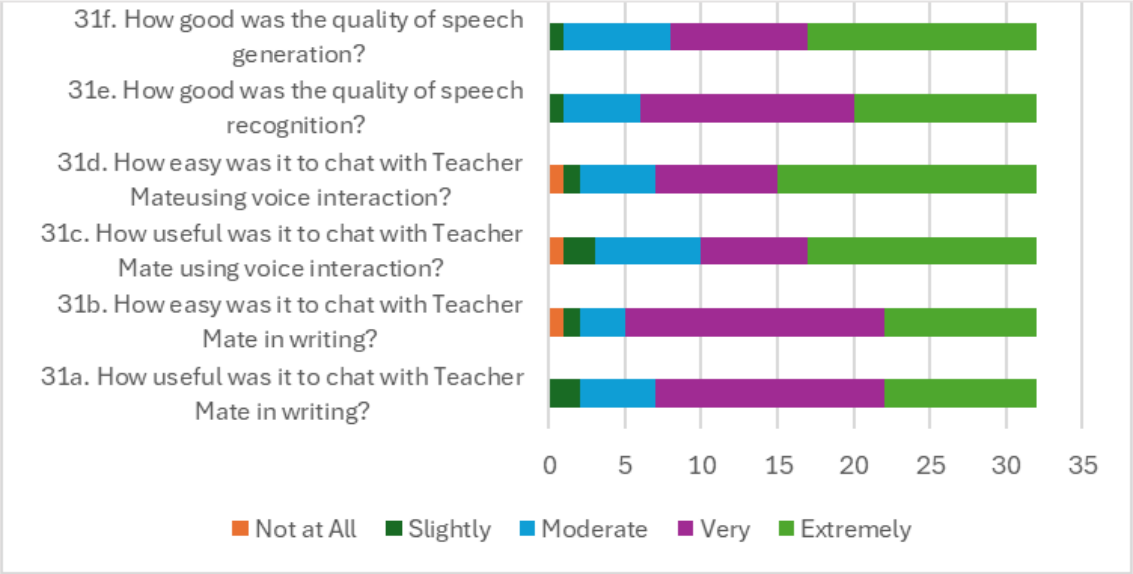

Figure 9: TM - Question 31 Results

4.1.4. Experience on using the Chat History functionality

This section presents the evaluation of the Chat History functionality of the TM conversational AI assistant, focusing on teachers' experiences and feedback. The questionnaire included seven key rating questions:

Co-funded by the European Union

- The accuracy of the short titles of previous chats.
- The ease of going through the Chat History.

A subset of questions was included aiming to evaluate the quality of the TM's responses in terms of accuracy, language, coherence, ethics and depth of explanation. Participants were given time to go through their own previous conversations with the TM, read them carefully and provide their feedback on these quality criteria. Specific instructions were given to pay attention to the distinction between the general chat conversations, conducted using the general LLM, and textbook specific conversations, which used the textbook as a source of the LLM's response, so the effectiveness of RAG is evaluated.

- The overall accuracy of the information provided in TM's responses.
- The correctness of the language used in TM's responses.
- The coherence of TM's responses.
- The ethical considerations in TM's responses.
- The depth of the responses in terms of insightful perspectives or additional resources for deeper understanding.

The subsequent analysis will highlight overall user satisfaction and areas for potential improvement of the Chat History component, so it supports effective review and utilization of past interactions with the AI assistant.

Accuracy of Short Titles of Previous Chats (34a)

The survey responses for the accuracy of the short titles of previous chats reveal positive feedback. Out of 32 respondents, the majority (25) rated the accuracy as either "Very" accurate or "Extremely" accurate. Specifically, 19 respondents rated it "Very" accurate, and 6 rated it "Extremely" accurate. A smaller number of respondents found the accuracy to be only moderate, with 5 individuals selecting "Moderate." Only 2 respondents rated it slightly accurate, and none rated it not at all accurate, suggesting a generally favorable view of this feature.

Ease of Going Through the Chat History (34b)

Responses regarding the ease of going through the Chat History show a significant lean towards positive feedback. Out of the 32 respondents, 24 found the process either "Very" easy or "Extremely" easy. Specifically, 16 respondents rated it as "Very" easy, and 8 rated it "Extremely" easy. This overwhelming positive feedback highlights the intuitive and user-friendly nature of this feature. A smaller number of respondents, 6, rated the ease of use as "Moderate," indicating that some users might encounter occasional difficulties but generally find the tool manageable. Only 2 respondents rated it slightly easy, and none rated it not at all easy, further underscoring the overall positive reception of the tool's usability.

Co-funded by
the European Union

Accuracy of Information in TM's Responses (34c)

The survey responses regarding the accuracy of the information provided in TM's responses show a significant lean towards positive feedback. Out of the 32 respondents, 22 found the information to be either "Very" accurate or "Extremely" accurate. Specifically, 14 respondents rated it as "Very" accurate, and 8 rated it "Extremely" accurate. This indicates that users generally trust the accuracy of the information provided by TM. A smaller number of respondents, 6, rated the accuracy as "Moderate," indicating occasional inconsistencies or reservations. Only 3 respondents rated it negatively, with 1 marking "Not at All" accurate and 2 marking "Slightly" accurate, further underscoring the overall positive reception of the tool's accuracy.

Correctness of Language Used in TM's Responses (34d)

The survey responses regarding the correctness of the language used in TM's responses indicate positive feedback. Out of 32 respondents, the majority (22) rated the language as either "Very" correct or "Extremely" correct. Specifically, 15 respondents rated it "Very" correct, and 7 rated it "Extremely" correct. This high level of satisfaction suggests that the language used by TM is generally considered accurate and appropriate. A smaller number of respondents found the language to be only moderately correct, with 5 individuals selecting "Moderate." Only 5 respondents rated it negatively, with 2 marking "Not at All" correct and 3 marking "Slightly" correct, indicating some areas for improvement in language use.

Coherence of TM's Responses (34e)

The survey responses regarding the coherence of TM's responses are largely positive. Out of the 32 respondents, 20 found the responses either "Very" coherent or "Extremely" coherent. Specifically, 15 respondents rated the responses as "Very" coherent, and 5 rated them as "Extremely" coherent. This indicates a high level of satisfaction with the coherence of TM's responses. A notable number of respondents found the coherence to be only moderate, with 9 individuals selecting "Moderate." Only 3 respondents rated the coherence negatively, with 1 marking "Not at All" coherent and 2 marking "Slightly" coherent, suggesting some areas for improvement in maintaining coherence in responses.

Ethical Nature of TM's Responses (34f)

The survey responses regarding the ethical nature of TM's responses are overwhelmingly positive. Out of 32 respondents, the majority (27) rated the responses as either "Very" ethical or "Extremely" ethical. Specifically, 20 respondents rated the responses as "Very" ethical, and 7 rated them as "Extremely" ethical. This suggests a strong level of trust in the ethical standards maintained by TM. A smaller number of respondents found the responses to be only moderately ethical, with 5 individuals selecting "Moderate." None of the respondents rated the responses as slightly or not at all ethical, indicating a very positive perception of TM's ethical considerations.

Co-funded by
the European Union

Depth of Responses for Deeper Understanding (34g)

The survey responses regarding the depth of TM's responses in terms of providing insightful perspectives or additional resources for deeper understanding are positive. Out of 32 respondents, the majority (24) rated the depth as either "Very" deep or "Extremely" deep. Specifically, 21 respondents rated the depth as "Very" deep, and 3 rated it as "Extremely" deep. This indicates a high level of satisfaction with the depth of insights and resources provided. A smaller number of respondents found the depth to be only moderate, with 6 individuals selecting "Moderate." Only 2 respondents rated the depth negatively, marking it "Not at All" deep, indicating overall positive feedback but highlighting some areas for further enhancement in providing deeper insights.

In summary, the data suggests a strong positive reception of the various aspects of TM's chat functionalities and response quality evaluated in the survey. Most respondents find it easy to navigate the Chat History, consider the information and language used in responses to be accurate and correct, and view the responses as coherent and ethical. Additionally, most respondents appreciate the depth of the responses in terms of providing insightful perspectives and additional resources (Figure 10).

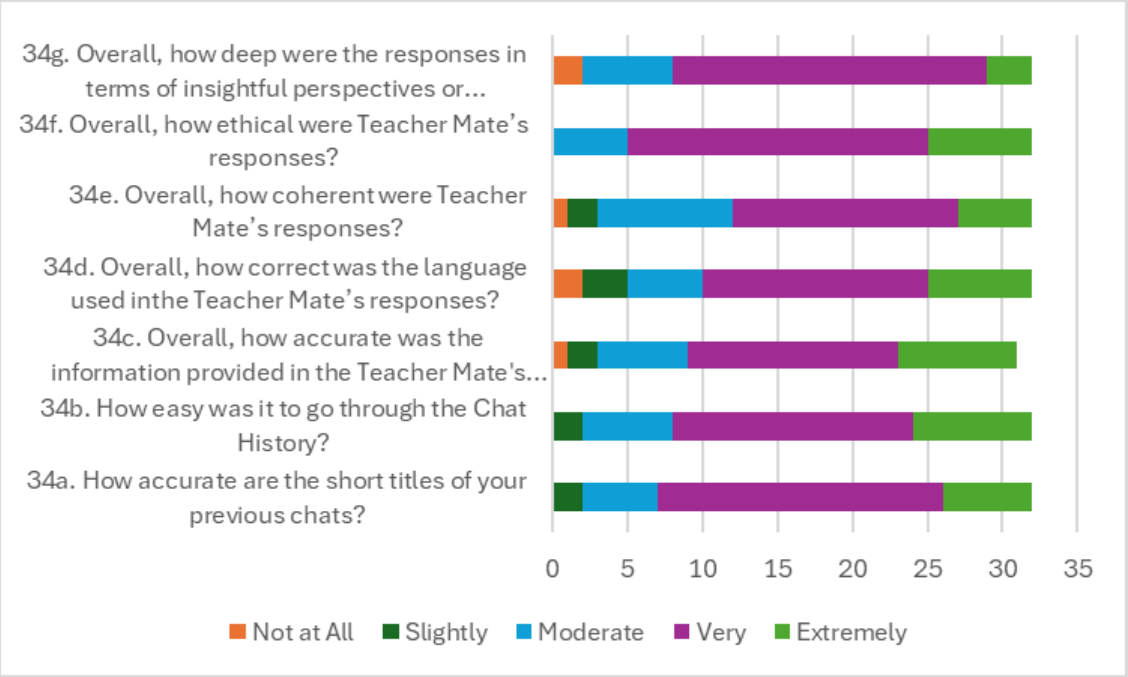

Figure 10: TM - Question 34 Results

Co-funded by the European Union

General Chat vs Text-book specific model (36)

The survey responses regarding the preference for the Teacher Mate's response generation models show a notable distribution of preferences. Out of the 32 responses, the "General Chat" model was slightly more preferred, with 14 respondents (43.75%) favoring it.

The "Text-book specific" model was preferred by 13 respondents (40.63%). Additionally, 5 respondents (15.63%) indicated no preference between the two models. This suggests a near balance in the effectiveness perceived by users between the general and specific response generation models, with a slight edge for the "General Chat" model.

In summary, the data reveals a nuanced preference pattern for the Teacher Mate's response generation capabilities. While the "General Chat" model holds a marginal lead in overall preference, the "Text-book Specific" model is also highly regarded, nearly matching the general model's popularity. The equal preference in Ireland and the exclusive preference for the "Text-book Specific" model in Sweden highlight the varying needs and expectations across different regions. The presence of a significant number of respondents with no specific preference indicates that both models are effectively meeting users' needs, though there remains a potential for further enhancement to cater to diverse user requirements comprehensively. (Figure 11).

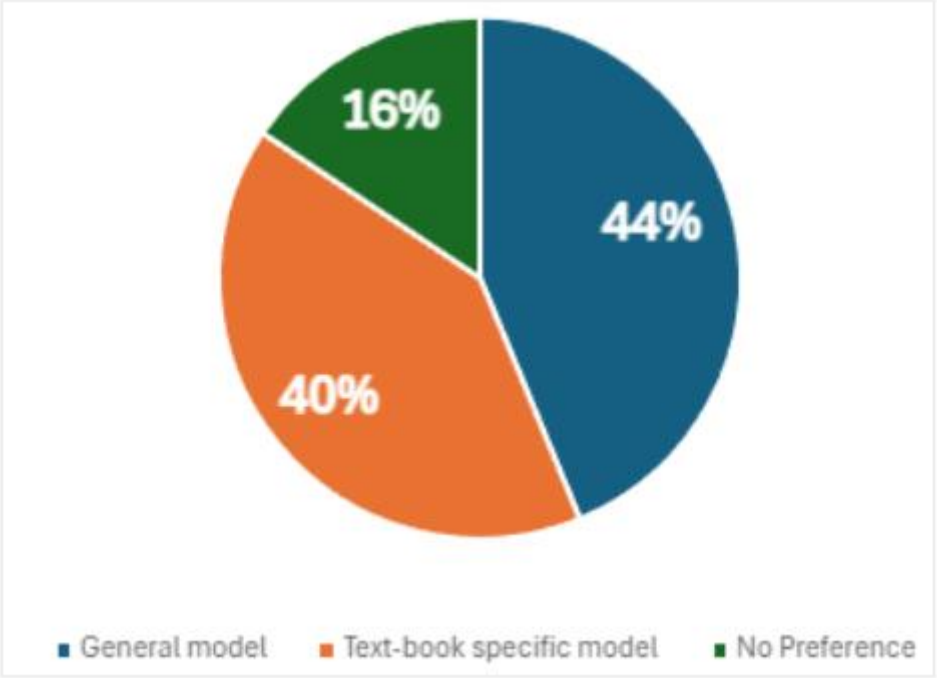

Figure 11: General Chat vs Text-book specific model

Co-funded by
the European Union

4.1.5. Evaluation of Teacher Mate's Usability and Features

This section explores the general experience of educators using the TM conversational AI assistant, focusing on a broad spectrum of user interactions and satisfaction metrics. The questionnaire included questions designed to capture detailed feedback on various aspects of the tool, including its graphical user interface, ease of use, integration of features, and overall effectiveness in supporting teaching practices.

More specifically, the question on Graphical User Interface (**Q41a**) evaluates users' satisfaction with TM's visual appeal and layout, crucial for user accessibility and engagement.

Question on frequency of Use (**Q41b**) determine teachers' willingness to use TM frequently, indicating its perceived value and utility in their teaching practice.

Question about the application's perceived complexity (**Q41c**) determines whether users find TM unnecessarily complex, balancing advanced functionalities with user-friendly design.

Questions on ease of navigation (**Q41d**), ease of use (**Q41f**), command Operation (**Q41h**) assesses how straightforward users find navigating and using TM's features, essential for consistent adoption and integration into daily practices.

Questions on feature Integration (**Q41e**) and complementarity with other resources (**Q41m**) examine how well TM's features complement existing teaching tools and practices, enhancing instructional approaches.

Question on clarity of interaction (**Q41g**) evaluates the understandability of interactions with TM, ensuring effective communication.

Questions about integration of the TM in daily teaching practice (**Q41i**) and classroom management effectiveness (**Q41l**) assess TM's usefulness in teaching scenarios, reducing workload and improving effectiveness.

Question **Q41j** reflects on whether TM empowers teachers to become more efficient with AI, enhancing teaching practices.

Questions on interest in use (**Q41k**) and safety of use (**Q41n**) measures user engagement and confidence in TM's safety and reliability.

Questions **Q41o** and **Q41p** evaluate TM's impact on subject knowledge enrichment and teaching quality, respectively.

Questions on resource Availability (**Q41q**) and knowledge sufficiency (**Q41r**) assesses access to necessary resources and teachers' comfort level with using TM.

Co-funded by
the European Union

Finally, **Q41s** and **Q41t** reflect the level of user independence in using TM and long-term commitment to the tool.

Quantitative results for each one of the above questions are presented below, following their order in the questionnaire"

41a. I like the Graphical User Interface of TM

The survey responses indicate a positive reception towards the graphical user interface (GUI) of TM. Out of 32 respondents, the majority (27) rated their experience as either "Agree" or "Strongly Agree." Specifically, 16 respondents marked "Agree," and 11 marked "Strongly Agree." This high level of satisfaction suggests that the GUI of TM is well-received by users. A smaller number of respondents were undecided, with 3 selecting "Undecided." Only 2 respondents marked "Disagree," and none rated it as "Strongly Disagree," indicating minimal dissatisfaction with the GUI.

41b. I think that I would like to use TM frequently

Responses regarding the frequency of using TM show a significant lean towards positive feedback. Out of the 32 respondents, 24 expressed a desire to use TM frequently, with 13 marking "Agree" and 11 marking "Strongly Agree." This suggests a high level of interest in regularly using TM. A smaller group, 5 respondents, were undecided. Only 1 respondent disagreed, and 2 respondents strongly disagreed, indicating a generally favorable interest in frequent use.

41c. I find the TM tools unnecessarily complex

The survey responses for the complexity of TM tools indicate mixed feedback. Out of 32 respondents, 24 rated the tools as either "Disagree" or "Strongly Disagree" in terms of unnecessary complexity. Specifically, 19 respondents marked "Disagree," and 5 marked "Strongly Disagree," suggesting that a majority do not find the tools complex. A smaller number of respondents, 5, were undecided. Only 2 respondents agreed, and 1 strongly agreed, indicating that complexity is not a significant issue for most users.

41d. I think TM is easy to navigate

The ease of navigation within TM received positive feedback. Out of 32 respondents, 26 rated it as either "Agree" or "Strongly Agree." Specifically, 17 respondents marked "Agree," and 9 marked "Strongly Agree." This suggests that users find TM easy to navigate. A smaller group, 3 respondents, were undecided. Only 2 respondents disagreed, and none strongly disagreed, indicating minimal issues with navigation.

Co-funded by
the European Union

41e. I find the features of TM are well integrated

The integration of features in TM received generally positive feedback. Out of 32 respondents, 21 rated the integration as either "Agree" or "Strongly Agree." Specifically, 18 respondents marked "Agree," and 3 marked "Strongly Agree." This indicates that users perceive the features as well-integrated. A notable number of respondents, 8, were undecided. Only 2 respondents disagreed, and 1 strongly disagreed, suggesting some room for improvement in feature integration.

41f. I find TM easy to use

The ease of use of TM received highly positive feedback. Out of 32 respondents, 29 rated it as either "Agree" or "Strongly Agree." Specifically, 18 respondents marked "Agree," and 11 marked "Strongly Agree." This indicates a high level of satisfaction with the ease of use. Only 1 respondent was undecided, and 1 strongly disagreed, with none marking "Disagree," indicating a generally positive user experience.

41g. My interaction with TM was clear and understandable

The clarity and understandability of interactions with TM received positive feedback. Out of 32 respondents, 26 rated their experience as either "Agree" or "Strongly Agree." Specifically, 16 respondents marked "Agree," and 10 marked "Strongly Agree." This suggests that users find interactions with TM clear and understandable. A smaller number of respondents, 2, were undecided. Only 3 respondents disagreed, and 1 strongly disagreed, indicating some areas for improvement in clarity.

41h. I found it easy to operate the various commands in TM

The ease of operating various commands in TM received highly positive feedback. Out of 32 respondents, 30 rated it as either "Agree" or "Strongly Agree." Specifically, 20 respondents marked "Agree," and 10 marked "Strongly Agree." This indicates a high level of satisfaction with the ease of operating commands. Only 1 respondent was undecided, and 1 marked "Slightly," with none marking "Disagree" or "Not at All," indicating a generally positive user experience.

41i. I think that TM will be useful in my daily teaching practice

The perceived usefulness of TM in daily teaching practice received positive feedback. Out of 32 respondents, 25 rated it as either "Agree" or "Strongly Agree." Specifically, 16 respondents marked "Agree," and 9 marked "Strongly Agree." This suggests that users see TM as a valuable tool in their daily teaching practice. A smaller number of respondents, 4, were undecided. Only 2 respondents disagreed, and none strongly disagreed, indicating minimal dissatisfaction.

Co-funded by
the European Union

41j. I think that TM can empower me as an efficient user of AI

The survey responses regarding the empowerment through TM as an efficient user of AI are largely positive. Out of the 32 respondents, 25 rated it as either "Agree" or "Strongly Agree." Specifically, 18 respondents marked "Agree," and 7 marked "Strongly Agree." This indicates a high level of perceived empowerment. A smaller number of respondents, 4, were undecided. Only 2 respondents disagreed, and none strongly disagreed, suggesting minimal dissatisfaction.

41k. I find the experience of working with TM interesting

The experience of working with TM was rated positively by most respondents. Out of 32 respondents, 30 rated it as either "Agree" or "Strongly Agree." Specifically, 15 respondents marked "Agree," and 15 marked "Strongly Agree." This suggests that users find the experience engaging and interesting. Only 1 respondent was undecided, and 1 marked "Slightly," with none marking "Disagree" or "Not at All," indicating a generally positive experience.

41l. I find TM effective for classroom management

The effectiveness of TM for classroom management received generally positive feedback. Out of 32 respondents, 21 rated it as either "Agree" or "Strongly Agree." Specifically, 10 respondents marked "Agree," and 11 marked "Strongly Agree." This indicates that users find TM effective for managing classrooms. A notable number of respondents, 7, were undecided. Only 3 respondents disagreed, and 2 strongly disagreed, suggesting some room for improvement in classroom management effectiveness.

41m. TM is complementary with other resources that I use

The complementarity of TM with other resources received generally positive feedback. Out of 32 respondents, 21 rated it as either "Agree" or "Strongly Agree." Specifically, 11 respondents marked "Agree," and 10 marked "Strongly Agree." This suggests that users find TM to work well alongside other resources. A notable number of respondents, 7, were undecided. Only 2 respondents disagreed, and 2 strongly disagreed, indicating some room for improvement in complementarity with other resources.

41n. I find the use of TM safe

The safety of using TM received highly positive feedback. Out of 32 respondents, 29 rated it as either "Agree" or "Strongly Agree." Specifically, 14 respondents marked "Agree," and 15 marked "Strongly Agree." This indicates a high level of trust in the safety of using TM. Only 1 respondent was undecided, and 1 marked "Slightly," with none marking "Disagree" or "Not at All," indicating a generally positive perception of safety.

Co-funded by
the European Union

41o. Using TM enables me to know more about the subject that I am teaching

The knowledge enhancement through TM received generally positive feedback. Out of 32 respondents, 19 rated it as either "Agree" or "Strongly Agree." Specifically, 15 respondents marked "Agree," and 4 marked "Strongly Agree." This suggests that users find TM helpful in gaining more knowledge about their teaching subjects. A notable number of respondents, 8, were undecided. Only 2 respondents disagreed, and 2 strongly disagreed, indicating some room for improvement in knowledge enhancement.

41p. By using TM, the quality of my everyday teaching tasks will be improved

The improvement in teaching quality through TM received highly positive feedback. Out of 32 respondents, 25 rated it as either "Agree" or "Strongly Agree." Specifically, 20 respondents marked "Agree," and 3 marked "Strongly Agree." This indicates a high level of perceived improvement in teaching quality. Only 4 respondents were undecided, and 2 marked "Slightly," with none marking "Disagree" or "Not at All," suggesting overall positive feedback.

41q. I have the necessary resources (PC, Mobile device, Internet Connection, Data) to use TM

The availability of necessary resources for using TM received overwhelmingly positive feedback. Out of 32 respondents, 32 rated it as either "Agree" or "Strongly Agree." Specifically, 10 respondents marked "Agree," and 22 marked "Strongly Agree." This indicates that users feel well-equipped with the necessary resources. None of the respondents were undecided or marked any negative options, indicating complete satisfaction with resource availability.

41r. I have the necessary knowledge to use TM

The survey responses regarding the necessary knowledge to use TM are highly positive. Out of the 32 respondents, 28 rated it as either "Agree" or "Strongly Agree." Specifically, 13 respondents marked "Agree," and 15 marked "Strongly Agree." This suggests that users feel confident in their knowledge to use TM effectively. Only 4 respondents were undecided, with none marking any negative options, indicating a high level of confidence in their knowledge.

41s. I could use TM, even if there was no one around to support me

The survey responses regarding the ability to use TM independently are highly positive. Out of 32 respondents, 28 rated it as either "Agree" or "Strongly Agree." Specifically, 13 respondents marked "Agree," and 15 marked "Strongly Agree." This suggests that users feel capable of using TM without additional support. Only 4 respondents were undecided, with none marking any negative options, indicating a high level of confidence in independent use.

Co-funded by
the European Union

41t. I intend to use TM in the future

The intention to use TM in the future received highly positive feedback. Out of 32 respondents, 27 rated it as either "Agree" or "Strongly Agree." Specifically, 14 respondents marked "Agree," and 13 marked "Strongly Agree." This indicates a strong intention among users to continue using TM. Only 2 respondents were undecided, and 1 marked "Slightly," with none marking any negative options, suggesting a high level of future engagement with TM.

The data from the survey reveals a **strong positive reception** towards various aspects of TM. Most respondents find the **graphical user interface** (GUI) appealing, which indicates that the design and layout are user-friendly and visually pleasing. Additionally, a significant number of respondents expressed a desire to use TM frequently, showing enthusiasm to incorporate TM into their routine. This **frequent usage** is likely driven by the perceived ease of use, as 31 respondents found it easy to navigate and operate. Such ease of use is crucial for adoption, as it lowers the barrier to entry for new users and ensures that existing users can use the tool effectively without extensive training.

However, there is a noticeable divergence in responses regarding the **complexity** of TM's tools. While the majority of respondents indicated satisfaction with the tool's simplicity, a smaller number undecided or found the tools complex. This suggests that there is a part of end users who might benefit from additional support, training or simplification of certain features.

The **coherence, clarity, and ethical nature** of TM's responses were rated highly, with most respondents finding the responses clear, coherent, and ethical. This high level of satisfaction is critical as it reflects the tool's ability to provide reliable and comprehensible assistance, fostering trust among users.

The **integration** of TM with other resources and its **effectiveness for classroom management** also received positive feedback. Most respondents agreed that TM complements other resources they use, which is important for seamless integration into existing workflows. The tool's effectiveness in classroom management further highlights its utility in an educational setting, making it an asset for teachers.

Respondents also indicated that TM **enhances their knowledge** about the subjects they teach and **improves the quality** of their teaching tasks. This dual benefit underscores the tool's educational value. It not only aids teachers in delivering better education but also supports their professional development.

Confidence in the availability of **necessary resources and knowledge** to use TM is overwhelmingly positive. All respondents felt they had the necessary resources, and the vast majority felt confident in their knowledge to use the tool. This confidence likely contributes to the high levels of reported ease of use and satisfaction.

Co-funded by
the European Union

Finally, the **intention to use** TM in the future is strong among respondents, with the vast majority indicating they plan to continue using the tool. This strong future intention, combined with the high levels of current satisfaction and ease of use, reflects high potential for sustained adoption and success of TM.

In conclusion, the survey data paints a comprehensive picture of TM as a highly valued, easy-to-use tool that effectively supports teachers in their daily tasks, enhances their subject knowledge, and integrates well with other resources. While there are areas for potential improvement, particularly in simplifying certain tools and ensuring coherence of responses, the overall positive reception suggests that TM is well-positioned to continue supporting educators effectively (Figure 12).

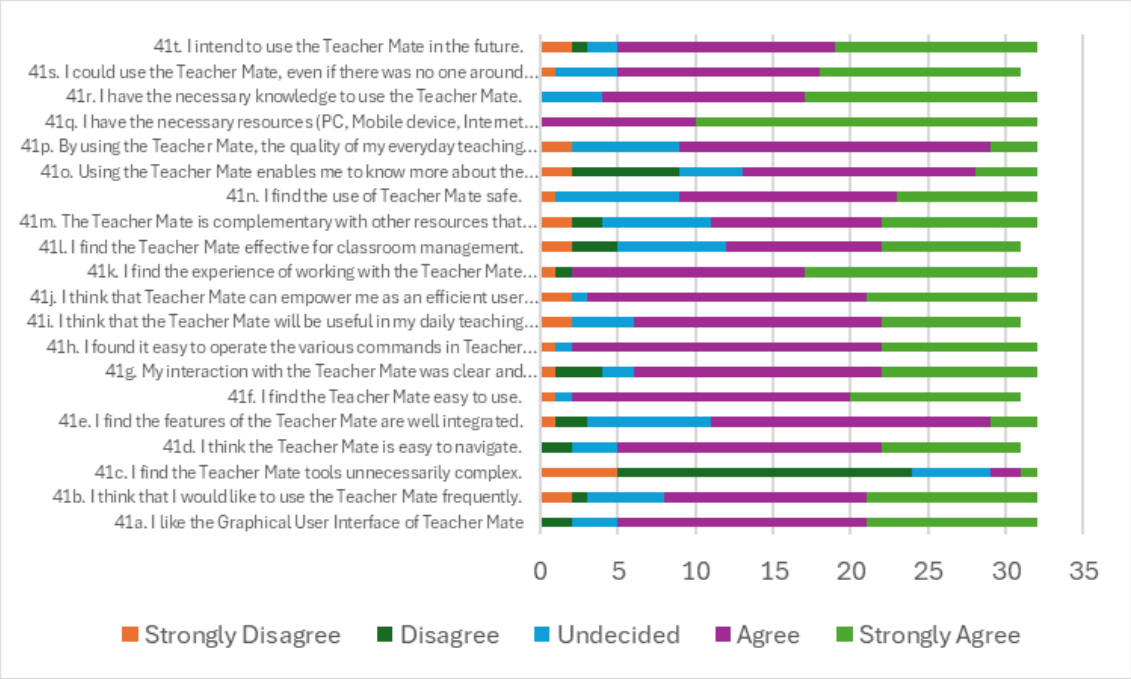

Figure 12: TM - Question 41 Results

4.2. Qualitative Analysis of Teacher Mate

This section presents the qualitative feedback collected from teachers through open-ended questions. These questions aimed to gather educators' suggestions, recommendations for improvement, and insights into TM's tools and features' perceived usability and technology acceptance. For each component of the TM, the participants were asked:

a) to suggest any other parameters for prompt creation or functionalities that could be added to improve the tools' effectiveness, and
b) to share any other observations -based on their hands-on experience- of the tools' performance that will guide the improvement of tools in terms of usability and overall effectiveness.

Co-funded by
the European Union

The answers to these open-ended questions complement the quantitative analysis presented in section 4.1. Subsections of 4.2 follow the order of the 4.1 subsections, which is aligned to the questionnaire structure.

4.2.1. Evaluation of the Teacher Mate tools

4.2.1.1. Lesson Plan Creation

Suggested Additional Parameters for the Lesson Plan Creation tool

Responses on suggested additional parameters for the Lesson Plan Tool provided interesting insights. One user expressed a strong critique, noting that the **AI model is insufficiently trained**, particularly in **Greek**. This user suggested that, in its current state, the tool is not useful. This feedback underscores the need for better training of the AI in different languages and possibly the inclusion of parameters that address language proficiency and localization.

Another user proposed the inclusion of parameters that allow for **multiple sources of knowledge** to be incorporated into the lesson plan. This suggests a need for the tool to support diverse educational resources and provide a more holistic approach to student assessment. Including the degree of difficulty for subjects is also suggested. This would help tailor lesson plans to different student abilities and ensure that the content is appropriate for the targeted age and skill level.

Feedback included the suggestion to incorporate **curriculum standards**, such as the 5E instructional model, and to allow for lesson plans to be exported to word documents for easier sharing and editing. There is also a call for aligning the lesson plans with the national or local curriculum standards to ensure relevance and compliance.

Users highlighted the importance of being able to include **lesson goals** and new information beyond the standard textbook content. This aligns with modern educational practices that emphasize setting clear learning objectives and encouraging exploration beyond the textbook.

Several users emphasized the need for **detailed success and proficiency indicators**. These indicators are essential for measuring student progress and ensuring that the lesson objectives are met. Suggestions included specifying which indicators are covered in each lesson and incorporating detailed curriculums along with these indicators.

The feedback also indicated a need for **utilizing visual materials** and providing suggestions for alternative assignments, such as virtual reality experiences and museum visits.

There were suggestions for the tool to include options for teachers to state the **resources** available at their school and to specify the types of activities they want to generate, such as experiments, presentations, or field trips.

Observations and Comments for Improvement

Co-funded by
the European Union

One user requested **clearer, more detailed, and structured activities** in the lesson plan suggestions. This implies a need to apply more sophisticated prompt engineering methods so the AI provides more specific and actionable recommendations rather than general pedagogical methods.

Some feedback indicated that the lesson plan components were not always **relevant to the unit requested**. Ensuring that the tool accurately aligns the lesson content with the specified unit of the textbook is crucial for its effectiveness. A comment highlighted the need for the tool to be able to **stop when the conversation goes off track**.

Several users suggested adding **dropdown menus** within different fields of each form - corresponding to prompt parameters- for easier input. This would make the tool more user-friendly and efficient. For instance, dropdown menus or pre-filled fields with sample suggestions for activities, special education needs, and success indicators could facilitate the lesson plan creation process.

One user praised the **voice facility feature**, noting it as an excellent intervention. This feedback indicates the importance of multimodal input options. Additionally, suggestions for incorporating **collaboration features** for projects highlight the need for tools that support group work and student interaction.

Adding categories for **age levels**, **ability levels**, and **special needs** (e.g., EAL, visual/speech/hearing impairments) was suggested to ensure that lesson plans are inclusive and appropriately tailored to diverse student populations.

Table 32: TM Lesson Plan Creation – Summary of Qualitative Research Results

Areas of improvement – Suggested features
• **Improvement of language quality (particularly in Greek)** • **Parameters for teaching approach, complexity, execution time, and originality.** • **Support for multiple knowledge sources.** • **Include the degree of difficulty for subjects.** • **Alignment with curriculum standards (e.g., 5E instructional model).** • **Export lesson plans to Word documents.** • **Incorporate lesson goals and new information beyond textbooks.** • **Detailed success and proficiency indicators.** • **Use visual and interactive materials.** • **Include resource availability and activity types.** • **Clearer and more detailed responses.** • **Relevant unit-specific content.** • **Mechanism to stop off-track conversations.** • **Dropdown menus and improved UI.** • **Voice facility and collaboration features.** • **Age and ability level customization.**

Co-funded by
the European Union

4.2.1.2. Quiz Creation

Suggested Additional Parameters for the Create Quiz Tool

Users suggest including parameters that specify the **level of difficulty** of questions and the estimated **time** needed to complete the quiz. These additions would help tailor quizzes to different student abilities and ensure they are appropriately challenging.

There is a recommendation to account for students' **learning difficulties**. This could involve setting different time limits for answering questions and providing additional resources for individual study based on the student's answers. These features would make the quiz tool more adaptive and supportive of diverse learning needs.

A parameter for **individual assessment and feedback is suggested**, which would allow teachers to provide personalized feedback based on quiz performance. This feature is covered by the Create Test Tool, which is designed to support formative assessment practices.

Users emphasize the need for the ability to **add images, diagrams, and audiovisual material** to quizzes. This is particularly important for subjects like Biology and History, where visual aids are essential for understanding the content.

Incorporating success and proficiency indicators from **educational standards** (such as those from the Ministries of Education) is recommended. This would ensure that quizzes are aligned with official curriculum guidelines and help track student progress more effectively.

There are suggestions to improve the **format and structure of certain types of questions**, such as matching questions and multiple-choice questions. For example, adding extra options in multiple-choice questions to reflect exam paper guidelines and including exercises for chronological sequencing of events in History.

Additionally, differentiation options to cater to **different age groups and ability levels** were reported, to make the tool more inclusive and versatile.

Observations and Comments for Improvement

Users note that the tool currently exhibits **errors in syntax and terminology**, particularly in languages other than English, such as **Greek**. The linguistic accuracy of the TM's responses is limited by the LLM's capabilities, however, both SB's and TM's design allows support from various LLMs, which can be more accurate in less widely used languages, such as Greek and Swedish.

Some users reported that the quizzes generated **did not match the requested content** or were not helpful in creating relevant quizzes.

Co-funded by
the European Union

Feedback includes suggestions to improve the **user interface**, such as spreading out the questions on the screen to avoid clutter and adding dropdown menus for different quiz types (e.g., multiple-choice, open-ended, cloze procedures).

Users express a desire for the tool to create more interactive and **multi-modal quizzes**, which would include various types of questions and multimedia elements.

It is mentioned that the process of **saving conversations or quiz settings** is not clear. Providing a straightforward and transparent way to save and retrieve work would enhance the tool's usability.

Table 33: TM Quiz Creation - Qualitative Research Results

Areas of Improvement - Suggested Features
• **Parameters for question difficulty and completion time.** • **Consideration of learning difficulties.** • **Individual assessment and feedback options.** • **Inclusion of visual and audiovisual materials.** • **Success and proficiency indicators.** • **Improved question formats and structures.** • **Multimodal features and differentiation options.** • **Syntactic and terminological accuracy.** • **Relevance to requested content.** • **Support for visual materials.** • **User interface improvements.** • **Interactive and multi-modal quizzes.** • **Clear saving mechanisms.** • **Option to differentiate between test and practice questions.**

4.2.1.3. Presentation Creation

Suggested Additional Parameters for the Create Presentation Tool

Several users suggest making the presentation output **more structured**, with specified slide numbers, bullet points, and suggested images for each slide. This would guide users more effectively and provide a clear template to follow.

There is a call for additional interactive tools and the creation of a digital library with **images** to aid in creating presentations. Such resources would make the presentation tool more versatile and user-friendly, enabling users to access relevant materials easily. Users want the tool to develop presentations in more detail so that they are ready to be copied into PowerPoint or other presentation software. This includes having all the necessary elements,

Co-funded by
the European Union

such as images, videos, diagrams, and **detailed text**, to minimize the effort required by the user to finalize the presentation.

Feedback suggests that parameters could be more targeted to create presentations that **better match the specific unit or section** requested from a textbook. This would ensure that the content is relevant and aligned with user needs.

Users suggest referencing images and quizzes within the presentations and **linking** quizzes, presentations, and lesson plans. This integrated approach would enhance coherence and consistency across different teaching tools. Including **multimodal features** such as illustrations, graphs, photos, and audio aids is recommended, as well as adding details for all points on each slide.

Observations and Comments for Improvement

There are comments about the tool not **matching the content of the requested book** section and errors in **syntax and terminology** in **Greek**. Users suggest that the tool should allow for changes before finalizing the presentation. There is a desire for more **multimodal and inclusive options**, such as additional photos, music, and voices. Some users express that it would be incredible if the tool could directly produce a **PowerPoint or Google Slides** presentation.

Changing the words "**Course**" and "**Topics**" to "**Subject**" and "**Subject Area**" is suggested to make the tool more contextually appropriate for certain educational settings. Feedback indicates that users find it cumbersome to go back and **change previously selected options**, such as a book.

Table 34: TM Presentation Creation Tool - Qualitative Research Results

Areas of Improvement - Suggested Features
• **Better alignment of response to the textbook content.** • **Improved language quality (especially in Greek)** • **Structured slides with suggested content.** • **Interactive tools and digital library.** • **Detailed and ready-to-use presentations.** • **Targeted parameters for specific content.** • **Integration with quizzes and lesson plans.** • **Multimodal features.** • **Detailed slide content and visual aids.** • **Inclusion of non-Western sources.** • **Flexibility and customization.** • **Automated presentation creation.** • **Terminology adjustments.** • **Improved UI and workflow.** • **Speed and efficiency improvements.**

Co-funded by
the European Union

4.2.1.4. Grading

Suggested Additional Parameters for the Grading Tool

Users suggest incorporating parameters if **test duration** and **learning difficulties**. This would enable the tool to provide more personalized and fair assessments, taking into consideration the diverse needs of students.

Including parameters for the **distribution of marks** per different criteria is suggested. Additionally, users propose the inclusion of different **grading scales** and the ability to grade using multiple correct answers. This flexibility would cater to various grading systems and allow for a more comprehensive evaluation of student responses.

Incorporating rubrics for grading open-ended questions is highly recommended. Rubrics provide clear criteria for assessment, ensuring consistency and transparency in grading subjective responses.

Adding parameters that allow for the **comparison of student performance across different units** would provide valuable insights into learning progress and areas that may require additional focus. These capabilities are largely covered by the Test Creation tool of the TM.

Users emphasize the importance of considering the **special needs of students** in the grading process. This includes accepting answers given in general conversation and not strictly adhering to textbook responses, accommodating various learning styles and needs.

Including age as a parameter is suggested to ensure that the grading tool provides age-appropriate responses and assessments. This would make the tool more relevant and useful across different educational levels.

Observations and Comments for Improvement

Feedback indicates a need for improvement in the **assessment criteria** and the **accuracy** of answers provided to students. Ensuring that the grading closely follows the textbook or specified curriculum would enhance the reliability of the tool.

Issues of user experience were highlighted, such us suggestions that the entire question should be visible without needing to open additional fields, and the input field for answers should be larger.

Including the **pages** where answers are found and providing responses in the appropriate language (e.g., **Greek**) is recommended, to make the tool more user-friendly and contextually appropriate.

Users suggest having the option to **upload grading guides or rubrics**, possibly as PDF files, and upload students' answers as photos.

Co-funded by
the European Union

Improving the user interface to make it more user-friendly and ensuring the tool generates information quickly are important aspects for enhancing usability.

There is a desire for the tool to handle **multiple-choice question formats** effectively and to grade based on **different educational levels** (e.g., Junior/Senior Primary, Junior/Senior Cycle). This would make the tool versatile and applicable across various educational contexts.

Finally, including **basic default settings** (e.g., 1-5 grading scale) and **additional options for customization** would make the tool more adaptable to different grading practices and preferences.

Table 35: TM Grading Tool - Qualitative Research Results

Areas of Improvement - Suggested Features
• **Parameters for response time and learning difficulties.** • **Correction of specific question types.** • **Improved GUI** • **Improved grading accuracy** • **Distribution of units.** • **Different grading scales and multiple answers.** • **Rubrics for open-ended questions.** • **Comparison between units.** • **Consideration of special needs.** • **Age appropriateness.** • **Improved assessment criteria.** • **Accurate and relevant grading.** • **Page and language inclusion.** • **Upload and customization options.** • **User interface and speed improvements.** • **Multiple-choice and general chat grading.** • **Clear saving mechanisms.** • **Differentiation between test and practice questions.**

4.2.1.5. Creating Teaching Material

Suggested Additional Parameters for the Create Teaching Material Tool

Users emphasize the importance of integrating various **multimedia**, such as images, drawings, shapes, videos, and sound, to support the creation of more engaging and interactive teaching materials.

Including parameters to specify the **type of teaching material** (e.g., concept maps, historical timelines, interactive maps) and the **age group** addressed would help tailor the content more effectively to the users' needs. Incorporating **multiple sources** of information is suggested to

Co-funded by
the European Union

enrich the teaching materials. This would ensure that the content is comprehensive and covers various perspectives.

Users recommend providing more **detailed explanations of the key elements** of the course. This would help educators understand and convey the essential aspects of the subject matter more clearly.

Providing **a list or some tips for selecting type of educational material** they want to create is suggested.

Including **suggestions for assessment** within the teaching materials would provide educators with ready-to-use evaluation tools, helping them measure student understanding and progress effectively.

Allowing for **more detailed analysis** of each point presented in the teaching materials is suggested. This would provide deeper insights and a more thorough understanding of the subject matter.

Observations and Comments for Improvement

Users note that the responses from the tool could be more targeted to ensure **relevance** to the requested content.

There are concerns about the **accuracy of language** and context. Ensuring that the tool uses the correct terminology and provides references in the appropriate language (e.g., English) is relevant to its effectiveness.

Utilizing **visual material** and providing suggestions for visual aids are highlighted as important aspects. Presenting each point with a **more detailed** analysis is recommended. This would help educators convey complex information in a more understandable and comprehensive manner.

Feedback indicates that the tool sometimes provides **inconsistent answers** to similar questions asked in different contexts.

One user suggests including **books or sources from various countries** to make the tool more internationally relevant.

Adding **tooltip text** over each parameter is suggested to help users understand the options and make more informed choices. This would enhance the user experience by providing additional guidance.

Including **blank fields or dropdown menus** to select the **class level** or **year grade** that the content is aimed at would make the tool more user-friendly and ensure that the materials are appropriate for the intended audience.

Co-funded by
the European Union

Linking the created materials back to the **curriculum** and providing differentiation options for **diverse learning needs** are recommended. This would ensure that the teaching materials are aligned with educational standards and cater to different student abilities. Clarifying the terminology for special needs to avoid confusion is suggested. For example, changing "Special Needs" to a term that better reflects the intended meaning, such as "Additional Support Needs," would make the tool more inclusive and accurate.

Table 36: Teacher Creating Teaching Material Tool - Qualitative Research Results

Areas of Improvement - Suggested Features
• **Multimedia integration.** • **Type of teaching material and age group specification.** • **Multiple sources of information.** • **Detailed explanation of key elements.** • **List of educational material types.** • **Assessment suggestions.** • **Detailed analysis of points.** • **Targeted responses.** • **Language and contextual accuracy.** • **Utilization and suggestions for visual material.** • **Consistency in responses.** • **International sources.** • **Tooltips for parameters.** • **Dropdown menus for class levels.** • **Curriculum relevance and differentiation options.** • **Clear terminology for special needs.**

4.2.1.6. Concept Exploration

Suggested Additional Parameters for the Concept Exploration Tool

Users suggest incorporating parameters that consider the **students' learning levels** and **curriculum**. This would allow the tool to provide more tailored and relevant information based on the classroom's specific needs and learning objectives.

Including **multiple sources of information** and assessing knowledge according to the students' learning levels and additional section elements is recommended. This would ensure that the concept exploration is comprehensive and appropriately challenging.

There is a need for the tool to be able to read previous answers and continue the dialogue seamlessly. Currently, it restarts the conversation, which disrupts the flow of learning.

Users recommend adding **archiving** functionality to the tool. This would allow educators to save, rename, and organize conversations for future reference, making it easier to track and revisit past discussions.

Co-funded by
the European Union

In line with previous suggestions, there is a call for access to **materials beyond textbooks**. This would provide a richer and more diverse set of resources for concept exploration. Adding detailed curricula and history textbooks from all grades to allow for horizontal concept searches across different books is suggested. This would enable comprehensive exploration of concepts across various educational materials. The ability to upload **photos** or written sources for processing and explanation would enhance the tool's utility.

Having a ready-made **dictionary of key terms** and historical figures from the curriculum would be useful. This would provide quick and easy access to essential information, facilitating better understanding and discussion. Including a **mind map** option to visualize concepts and spur discussion is recommended.

Incorporating **inter-curricular links** would allow for exploration of concepts across different subjects, providing a more integrated learning experience. Using explicit terms, especially for categories like "Special Needs," would make the tool more accessible and clearer for all users. There is a suggestion to link TM tools, e.g. having any questions generated under the Concept Exploration module sent to the Quiz module.

Observations and Comments for Improvement

Users note that the tool **struggles to locate and identify concepts** from specific **textbooks**, such as the Biology B' Gymnasium textbook. Ensuring that the tool can accurately reference and provide information from specified textbooks is crucial. There is feedback indicating that the tool only provides general information when users refer to specific books. Enhancing Retrieval Augmented Generation capabilities will improve the retrieval of detailed and specific information from the requested materials.

The tool should be able to **explain the differences** between similar concepts clearly. For example, it should differentiate between the result and consequence of a historical event accurately.

Improving the **speed of replies** is highlighted as an important aspect. One user suggests that while teachers might revert to the lesson plan tool for most concept explorations, the Concept Exploration tool could be particularly useful for **generating discussions**. The inclusion of comprehension questions at the end of the exploration is appreciated. This feature helps reinforce learning and assess understanding.

Table 37: TM Concept Exploration Tool - Qualitative Research Results

Areas of Improvement - Suggested Features
• **Improve interaction with textbook** • **General vs. textbook-specific information mismatch.** • **Parameters for learning levels and curricula.** • **Multiple sources and knowledge assessment.**

Co-funded by
the European Union

- **Seamless dialogue continuation.**
- **Archiving conversations.**
- **Access to materials beyond textbooks.**
- **Detailed curricula and horizontal concept search.**
- **Upload and process images or written sources.**
- **Ready-made dictionary of key terms and historical figures.**
- **Mind map option.**
- **Inter-curricular links.**
- **Explicit terminology.**
- **Integration with quiz module.**

4.2.1.7. Test Generation

Suggested Additional Parameters for the Create Test Tool

Users highlight the need for improving the **syntax and grammar** of questions and answers. This includes better wording of questions to ensure they are clear and grammatically correct.

Users also suggest **prompting modifications** to ensure that the same concept is not tested multiple times in different questions within the same test. This would help create more comprehensive and non-redundant tests.

Incorporating a **variety of exercise types**, such as True-False questions, is recommended. This would add diversity to the types of assessments available.

Adding **visual materials** such as maps, diagrams, or quotations from written or other sources would enhance the quality and engagement of the tests.

Users express the desire to **assign specific students** to the test and choose the source of information for the test, whether from the **textbook or a user-uploaded PDF**. This would provide more control and customization for the educator.

Observations and Comments for Improvement

There is a need for a **preview** option before submitting the test. Users should be able to see the final test format to ensure it meets their requirements before assigning it to students.

The tool should indicate the **page of the book** from which each question was taken. This would help educators verify the relevance and accuracy of the questions. Additionally, users suggest a **confirmation step** before the final assignment of the test to students.

The tool should recognize different **writing conventions** of answers in fill-in-the-blank or open-ended questions. This includes handling variations such as accents, uppercase letters, and **Greeklish** (a mix of **Greek** and English characters).

Co-funded by
the European Union

Providing **feedback** that influences teacher revision and exploring reply options would help educators refine their teaching strategies based on test outcomes.

Improving the user **interface** by highlighting active tests in green, done and assessed tests in red, and upcoming tests in gray would make it easier to manage and navigate through different tests. The option to hide already assessed tests is also suggested. Making the "Test Title" and "Description of Test's Subject" **more intuitive** would help users better understand and utilize these fields.

Table 38: TM Test Generation Tool - Qualitative Research Results

Areas of Improvement - Suggested Features
• **Parameters for syntax and grammar improvement.** • **Questions with parts and images.** • **Content coverage and redundancy avoidance.** • **Inaccurate and irrelevant grading.** • **Types of exercises.** • **Visual materials.** • **Assigning specific students and sources.** • **Preview and final test format.** • **Indicating source pages.** • **Recording information sources.** • **Editing and confirmation.** • **Feedback and revision.** • **Visual identification of test status.**

Visualisations of test performance data

Users suggest the ability to enter **additional information** to create a holistic view of each student's learning journey. Including **metrics** on class participation, deadlines, teamwork, and cooperative student spirit is recommended.

Users want to see the **progress** of each student individually, apart from overall class progress. This would help in identifying specific students' strengths and areas for improvement.

Visualizing gaps in relation to key skills is suggested, to target specific skills that need development. Additionally, users suggest assigning individual assignments to students, to enable more personalized learning and targeted interventions, and see statistics by exam/test, not just by student.

Providing **information on previous tests** would help track progress over time. Users suggest presenting data in **various formats**, such as pie charts, bar charts, and line graphs, so data interpretation is easier and more intuitive.

Co-funded by
the European Union

Including **school-based comparisons** to others using the tool at the same class levels is recommended. Isolating the progress in specific types of exercises or **comparing performance** across different chapters is suggested. Automatic notifications for teachers and students regarding any changes made would keep all stakeholders informed.

Ensuring **continuity of performance data** between years would help in tracking long-term progress. Allowing teachers to see all tests taken by a student with overall results in one click would improve efficiency in monitoring progress. Providing the ability to **edit student information** easily would enhance the tool's flexibility. A **visual indicator** (e.g., turning green) when marks are updated would confirm that the action has been completed successfully.

Table 39: : TM Visualizations - Qualitative Research Results

Areas of Improvement - Suggested Features
• **Comprehensive learning profile.** • **Class participation and teamwork metrics.** • **Individual student progress tracking.** • **Skill gap visualization.** • **Charts for individual students.** • **Connection with grading tool.** • **Statistics by exam/test.** • **Previous quizzes information.** • **Specific tasks assignment.** • **Visualization formats.** • **Comparison with other classes.** • **Class management options.**

4.2.2. Teacher Dashboard

One major concern is the current procedure for **creating classes**, which requires keeping student data in a file that cannot be modified or edited later. Although this approach ensures students anonymization and data privacy, this is perceived as a significant limitation.

Educators suggest that the dashboard should allow for the **import and export of information** from other systems, which would facilitate easier integration with existing data management tools. Furthermore, they emphasize the need to access and make **corrections to student accounts** at any time without having to download them repeatedly.

Providing options to **remove, add, and edit class names,** as well as add and remove students of a class was also suggested, to enhance usability and flexibility in managing classes.

The **linkage** between the test tool and automatic grading and feedback is another critical feature that users feel is lacking. Users highlight the need to access other tools within the dashboard to streamline the teaching process.

Co-funded by
the European Union

A **participation schedule** for each student, detailing their involvement in study activities and tests, was also suggested.

The need for **personal information** beyond class-related data is another area where improvements are suggested. Teachers would benefit from having a more holistic view of their students, which could include personal and academic information, helping them tailor their teaching strategies more effectively.

In terms of focus, the feedback indicates that the dashboard is currently **geared more towards secondary education**, which is the target end users' group of the AI4EDU project. Teachers are excited to see future developments that cater to **primary level education**, providing tools and functionalities that address the specific needs of younger students and their teachers.

Table 40: TM Dashboard - Qualitative Research Results

Areas of Improvement - Suggested Features
• **Streamlined process for creating and managing classes.** • **Integration with educational tools.** • **Flexible student data management.** • **Import and export information.** • **Persistent access to student codes.** • **Linkage between evaluation tool and automatic grading.** • **Participation schedule for each student.** • **Holistic view of students.** • **Save student usernames and passwords.** • **Ability to remove/add students from classes.** • **Voice options for teachers.**

4.2.3. Chat with Teacher Mate

The feedback gathered on the "Chat with TM" tool reveals several suggestions for enhancing its functionalities, alongside observations and opinions about the quality and accuracy of its responses. Users provided detailed insights, which can guide future improvements.

One of the primary suggestions involves the ability to **upload specific sites** and allow the tool to return answers based on those sites and **books** selected by users. This feature is perceived as highly beneficial in personalizing the learning experience and providing more targeted information. Additionally, users suggested that the tool should collect information by accessing **repositories of educational resources,** incorporate books from other countries and various sources like Amazon Kindle. This capability would help in curating appropriate content, aiding teachers in creating their own learning repositories.

Moreover, **archiving conversations** by topic, section, and class is suggested to enhance tool's usability, making it easier to revisit past interactions. Users also emphasized the need for

Co-funded by
the European Union

editing conversation titles and **grouping** conversations into categories they define, which would further enhance organizational efficiency.

Language support and **multimodal features** were also frequently mentioned. Users expressed the need for the tool to support multiple languages. Multimodal functionalities were suggested to enrich the learning experience, potentially including various formats like audio, video, and interactive content. There were suggestions for **language translation** options, which would be particularly useful in multicultural and multilingual classrooms.

Regarding user experience, many users found the tool to be very **teacher-friendly**, with features that cater to the needs of educators. Suggestions for improving the user experience included **linking questions** to previous answers to provide continuity in conversations.

When comparing the quality and accuracy of the **general** chat versus the **textbook-specific** chat, users provided **mixed feedback**. The **general chat** responses were often deemed **more accurate and thorough**, especially when the information from textbooks was limited or inaccurate. However, the **textbook**-specific chat was praised for offering **targeted and page-specific** answers, which can be particularly useful for teachers planning lessons.

Chat with the **textbooks** sometimes did not cover the necessary content comprehensively. Users noted that while many responses were useful, there were instances where the tool struggled with accuracy, sometimes providing incorrect or **off-topic** answers. This issue was particularly prominent with very specific questions that had multiple criteria, leading to confusion and less helpful responses. The **general chat**, pulling information from a broader range of sources, was often seen as **more suitable** for such needs.

Overall, users found the responses from the Chat with TM tool to be well-constructed and accurate, but there is room for improvement in precision and relevance. The feedback on the "Chat with TM" tool underscores the importance of personalization, differentiation between general and specific topics, language support, and enhanced educational resources.

Table 41: TM - Chat Tool - Qualitative Research Results

Areas of Improvement - Suggested Features
• **Enhanced UI.** • **Improved voice recognition.** • **Concise answers.** • **Interactive tools.** • **Better feedback mechanisms.**

4.2.4. Chat History

One of the main functionalities suggested by users is the ability to **sort answers** through tabs. This feature would help in organizing the chat history more efficiently, making it easier to navigate through past interactions.

Co-funded by
the European Union

Users also expressed the need to **change conversation titles** and to be able to go to specific answers for requesting clarifications. This would allow for more precise and focused follow-ups on previous discussions. Another significant recommendation was the **visibility of dates** in the chat history. Users mentioned that the date should be clearly visible to provide context for each conversation, which is essential for tracking the progression of topics over time.

Additionally, there was a call for **more distinct markers to differentiate responses** from the general chat and those based on textbooks. This distinction would help users understand the context and source of each response quickly.

The idea of incorporating a **search facility** within the chat history was widely supported. Users felt that having a robust search feature would allow them to retrieve specific information from their previous chats efficiently.

Exporting chat histories was another functionality that users deemed beneficial. The ability to export files would enable users to save and share important conversations, facilitating better collaboration and review.

Additionally, there was a suggestion for **slightly larger text** to improve readability, indicating a need for better accessibility features. The ability to **remove items** from the chat history was also mentioned, which would help users manage their chat logs better by eliminating irrelevant or outdated conversations.

An **alphabetical sorting** option was suggested to further enhance the organization of chat histories. There were some technical suggestions as well, such as exploring the use of advanced models like **ChatGPT-4o** to improve the quality of responses and overall functionality of the tool.

Users also wanted the chat history to be more integrated with other features, such as sending generated questions to a quiz module. This integration would create a more cohesive and comprehensive educational tool.

In terms of observations, users found it useful to have **a history saved**, as it allows them to review what has been previously done. This feature supports continuity and consistency in learning. Some users felt that they would be able to provide more detailed feedback after using the tool more extensively, suggesting that **ongoing use** could reveal additional insights and areas for improvement.

Overall, the feedback indicates a strong desire for better organization, searchability, and usability in the chat history functionality.

Table 42: TM Chat History - Qualitative Research Results

Areas of Improvement - Suggested Features
• **More precise chat titles.** • **Differentiation between general chats and chats with textbook**

Co-funded by
the European Union

- **Simplified navigation.**
- **Reliable and accurate responses.**
- **Language consistency.**
- **Search facility.**
- **Exporting chat histories.**
- **Larger text for better readability.**
- **Ability to remove items.**
- **Alphabetical sorting option.**
- **Integration with other features.**

4.2.5. General Acceptance and Usability feedback

The feedback on proposing similar tools or features and general observations about TM offers a wealth of suggestions and insights for improving its functionality and user experience. Users suggested several enhancements that could align TM with other advanced educational tools. One of the recurring suggestions was the integration of **multimodal** capabilities, that would greatly enrich the teaching materials and provide a more engaging learning experience for students. For subjects like History, adding different sources such as maps, diagrams, and images would make the tool more comprehensive and versatile.

Additionally, users recommended the feature to **upload personal books or PDFs** as resources. This functionality would allow teachers to utilize their own materials, ensuring that the content is highly relevant and tailored to their specific curriculum needs.

The importance of **archiving conversations** was also highlighted. Users felt that having a robust system to archive and retrieve past interactions would be invaluable for reference and continuity in teaching. Furthermore, the tool should provide clear indications of where information is **sourced** from—whether it is **general** or specific to the **textbooks**—especially in the creation of tests.

A significant suggestion was the ability to **export** all quizzes, tests, and class and student graphs to PDF. This feature would facilitate better record-keeping and sharing of academic progress with stakeholders such as parents and school administrators.

Additionally, incorporating the Success and Achievement Indicators, on the basis of each country's **educational standards**, would better align the AI's responses to the curriculum.

Another suggestion is the ability to **import** quizzes and tests created with other AI tools, allowing for seamless integration of previously developed materials.

In terms of general observations, users found TM to be **intuitive** and **well-connected**, appreciating its potential to support teaching. However, there were concerns about the **accuracy of page references** in questions based on textbooks, as they often did not correspond correctly to the actual textbook pages.

Users suggested adjusting parameter names to suit the **Irish** educational context.

Overall, users saw **enormous potential** in TM, describing it as an interesting, attractive, and well-designed resource. They appreciated its delivery and AI model, noting its suitability for

Co-funded by
the European Union

various languages and the possibility of extending its features to support visual and interactive inputs. While there were **no immediate drawbacks** mentioned, users expressed a desire to see more **targeted content** and functionalities that align with their specific teaching objectives.

5. System Data Analysis

Here we present a short analysis based on usage statistics collected by the system. Specifically, we collected statistics about the conversation length, duration and topics and here we present the results of this analysis, aggregated per country.

The system uptime was 90%. The downtime was due to a critical bug that was triaged before the Swedish pilot. The bug was successfully resolved during the pilot and after a few minutes the pilot was completed without any further obstructions.

5.1. Analysis of student interaction with Study Buddy

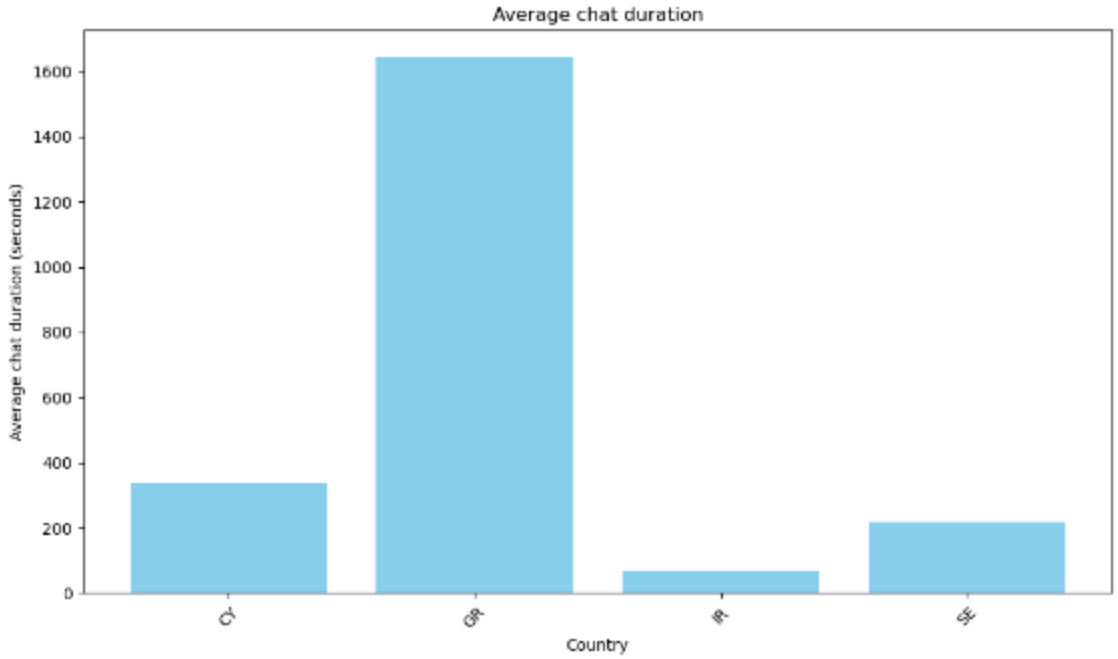

Figure 13: Average duration of student dialogs per country

In Fig. 13 we see the average duration of the student chats in seconds. Here we observe different usage patterns per country. In Greece we see the students engaged in long conversations (for an average of 26 minutes), while Irish students engaged in shorter conversations and mainly used the tools provided by StudyBuddy. Cypriot and Swedish students averaged about 5 minutes per conversations.

There may be multiple explanations for the different usage patterns to be considered. For example, the superior performance of the LLMs in English may facilitate shorter interactions

Co-funded by
the European Union

via tools, while in less popular languages follow-up queries may be more necessary. However, confounding factors during the pilot execution should be considered when interpreting these results, since the pilots were conducted in the dynamic school environment.

In Fig. 14 we also observe the average number of turns per conversation for each country. We see that the pattern follows the chat duration, further demonstrating the different usage patterns per country.

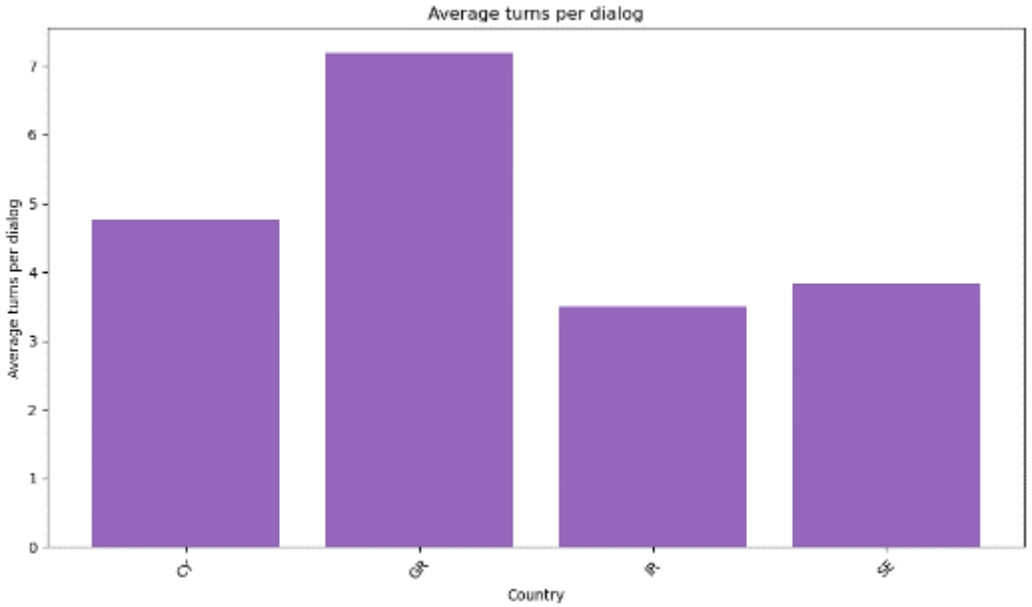

Figure 14: Average number of turns in student dialogs per country

Co-funded by
the European Union

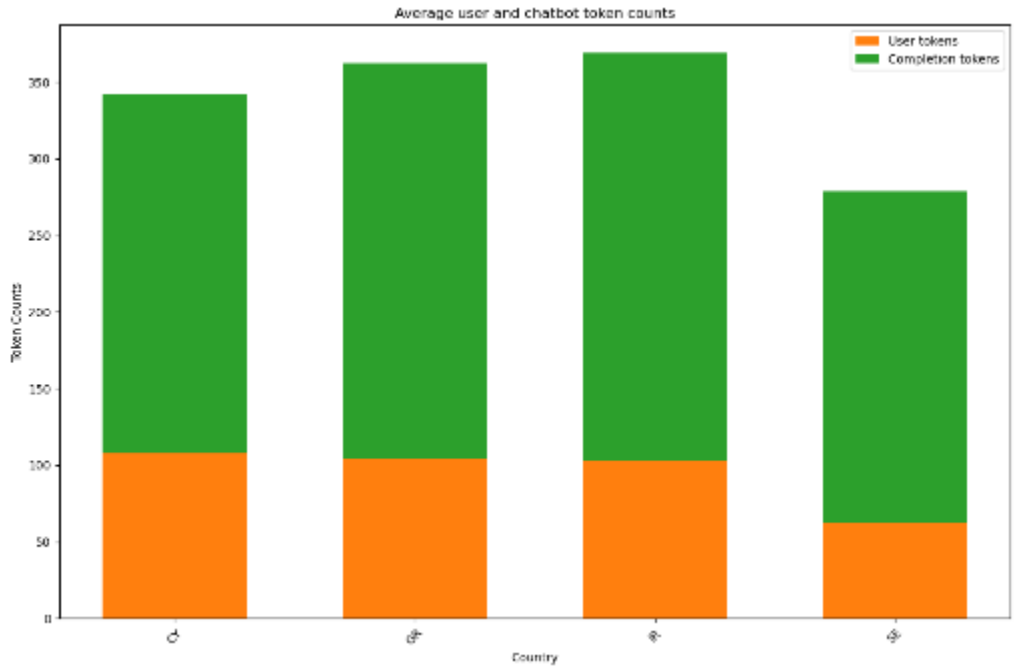

Figure 15: Average query (orange) and generated (green) tokens in student conversations per country

In Fig. 15 we see the average number of query tokens (orange) and generated tokens (green). We see that on average students wrote about 100 tokens for each conversation, while the average response consisted of 250 tokens.

Co-funded by
the European Union

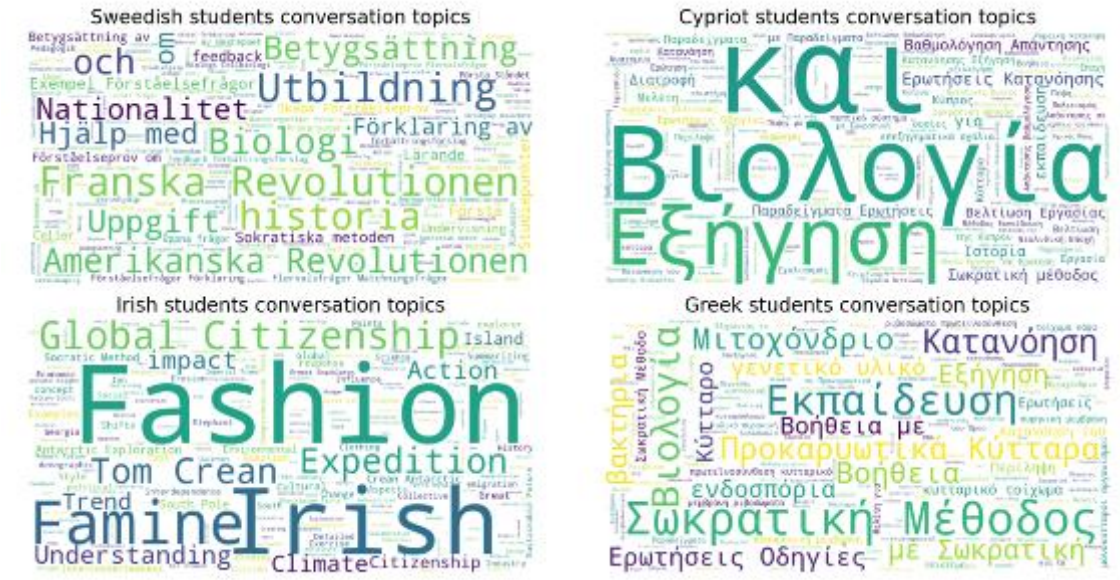

Figure 16: Common topics in student conversations

In Fig. 16 we see a wordcloud of the most common conversation topics per country. In Sweden we see that students explored social and historic topics, while in Cyprus and Greece Biology was more the focus, and term exploration using the Socratic method. In Ireland students focused on the topics of Irish history (e.g. the Great Famine) and ecological sustainability.

5.2. Analysis of teacher interaction with Teacher Mate

When analyzing system data collected from teacher interactions we see a different usage pattern. In Fig. 17 we see the average duration of teachers' interactions. Overall, teachers mainly utilized the tools offered by TeacherMate and opted for much shorter interactions. Irish teachers were especially engaged with the system. The use of tools primarily by the teachers is also indicated by the smaller number of conversation turns (2-3), as we see in Fig. 18.

Co-funded by
the European Union

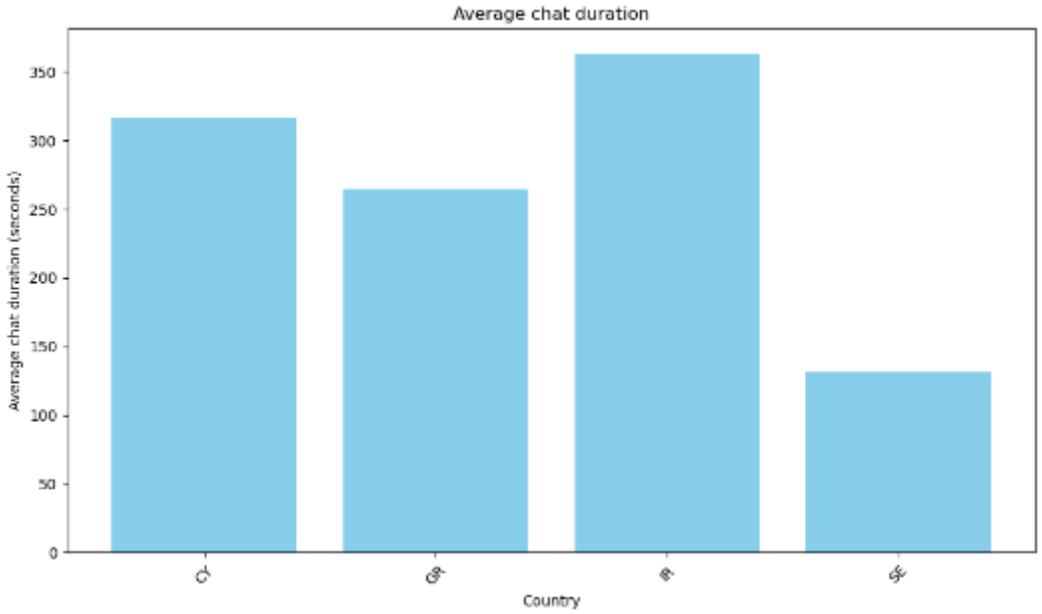

Figure 17: Average duration of teacher dialogs per country

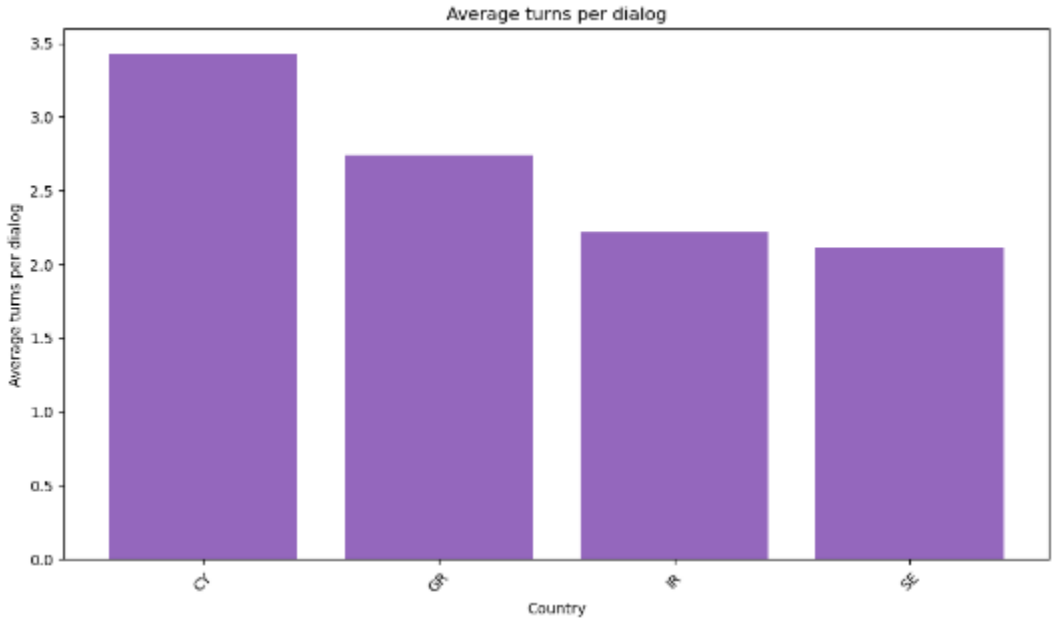

Figure 18: Average number of turns in teacher dialogs per country

Co-funded by
the European Union

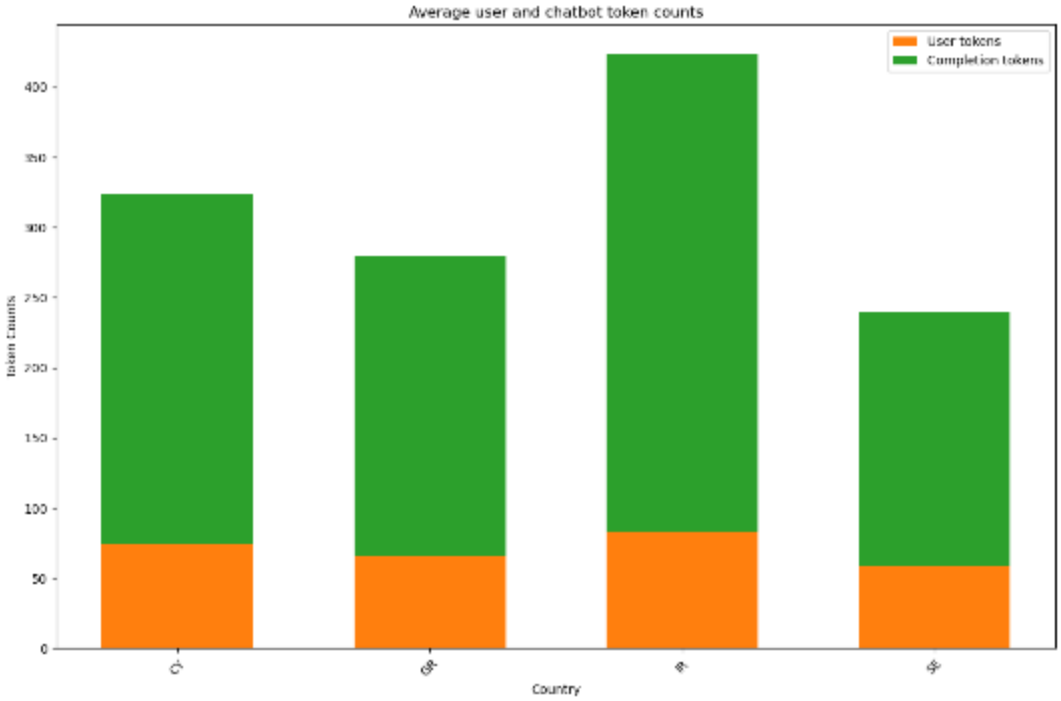

Figure 19: Average query (orange) and generated (green) tokens in teacher conversations per country

Finally, in Fig. 19 we see that the teachers opted for shorter prompts (60-70 tokens), while the average response length stayed around 200 tokens.

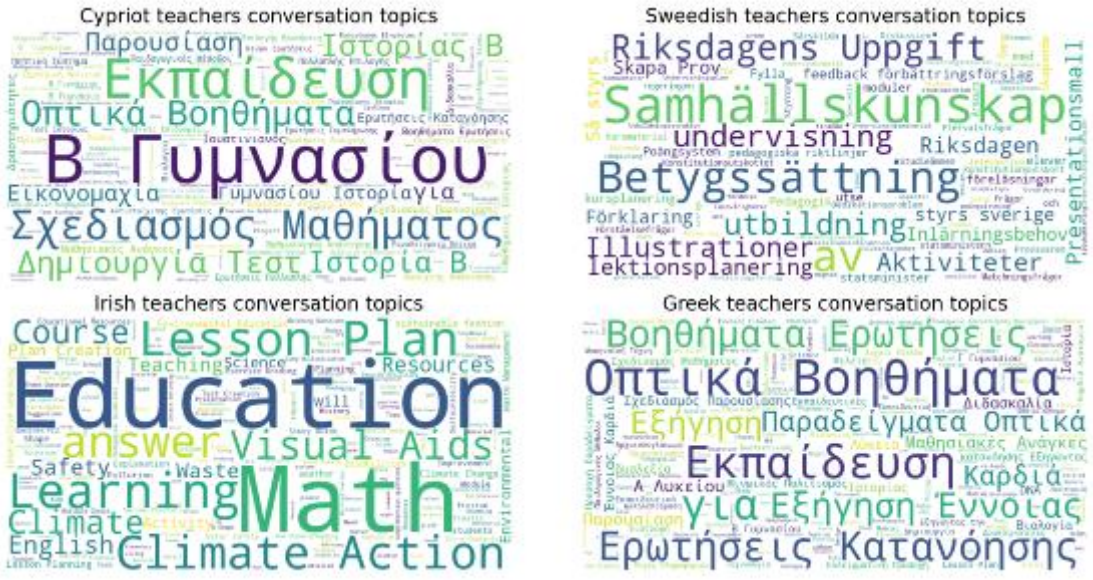

Figure 20: Common topics in teacher conversations

Co-funded by
the European Union

The more utilitarian use of the tool by the teachers is further showcased when we explore the wordclouds in Fig. 20. Here we observe that teachers focused on the creation of lesson content for their classes. Overall teachers focused on creating lesson plans, exploring better ways to explain terms to their students and creating rich content with visual aids.

6. Overall Conclusions & Recommendations

6.1. Summary of results

The quantitative and qualitative research results of the evaluation of the SB and TM tools provide a comprehensive understanding of their perceived usability and technology acceptance by students and teachers. The findings reveal diverse responses across different countries, highlighting both the strengths and areas for improvement for each tool.

The **SB** tool has several strengths and areas for improvement based on user feedback. Generally, the graphical user interface is well-received for its appealing and intuitive design, particularly noted in Cyprus and Ireland. Users in these regions also reported high satisfaction with the ease of use and navigation, indicating an effective and straightforward user experience. Additionally, users in Cyprus and Ireland highlighted positive learning impacts, reporting enhanced learning experiences and increased engagement with SB. Another strength is the perceived availability of necessary resources, as most users felt equipped to use SB effectively.

However, there are significant areas for improvement. One critical area is the **LLM's capabilities**. Users indicated a need for better handling of misspellings, improved contextual understanding, and more accurate responses, especially in non-English contexts. Another concern is the **complexity** and **usability** of the tool. While Cyprus and Ireland users were generally satisfied, feedback from Greece and Sweden suggested the need for **simplification** of tools and enhanced user guidance. Furthermore, to improve engagement and demonstrate practical benefits, particularly in Greece and Sweden, more effective **communication of SB's benefits** and applications is necessary. Lastly, enhancing **support** and **training** to boost user confidence and independence in using SB is identified as a crucial area for development.

TM has received positive feedback for several aspects, particularly its graphical user interface, which users in Ireland and Cyprus found visually appealing and well-laid-out. The usefulness of TM's tools for their educational routine was also praised, indicating that TM effectively supports teaching practice. Furthermore, users appreciated how TM empowers them and enhances their efficiency in using AI.

Nevertheless, there are areas that require significant improvements. One major area is **prompt engineering**. There is a need for more precise, context-aware, and curriculum-relevant prompts to improve the accuracy and relevance of AI responses. Additionally, the incorporation of **diverse educational resources**, **multimedia elements**, are suggested to enhance the learning experience.

Co-funded by
the European Union

User experience can be further improved by simplifying **navigation** and enhancing visual design. **Support and training** also need to be addressed comprehensively, providing tutorials and collect user feedback to refine AI responses and functionalities.

6.2. Conclusions & recommendations for the final phase of development

The feedback from the pilot results for SB and TM can be categorized into several key areas that will guide the next development phase, focusing on applications' improvements. These categories are based on insights from both quantitative and qualitative analyses, reflecting the experiences and suggestions of students and teachers.

The first key area is related to **User Interface and Experience**. Enhancements in navigation, additional platform features and front-end improvements towards ease of use were noted as essential regarding almost all of the SB and TM tools. Implementation of intuitive design elements like tooltips, prefilled fields of forms for prompt creation, menus, and labels, improvements in chat archiving, functionality of importing or exporting content, and more, were highlighted as suggested enhancements of the user experience across regions. Performance optimization to enable faster loading times and smoother interactions, was also suggested. Additionally, new functionalities that enhance communication between teachers and students, such as discussion forums and multilingual support were noted.

The second key area focuses on the **LLM's Capabilities**. Improving the AI's ability to handle misspellings, contextually understand queries, and provide accurate, relevant responses is critical, as noted by feedback from all regions. Additionally, handling multiple languages proficiently, particularly for non-English contexts, was emphasized, especially from Greece and Cyprus. Improved Retrieval-Augmented Generation (RAG) is also crucial to ensure that AI interactions are clear, concise, and directly address user queries in textbooks and educational resources. The linguistic accuracy of the TM's responses is limited by the capabilities of the LLM powering the conversational AI applications, SB and TM, which is currently Chat GPT 4.0. However, both SB's and TM's design allows integration of various LLMs, which can be more accurate in less widely used languages, such as Greek and Swedish. Recently, Athena RC has released the first Greek LLM, 'Meltemi', which shows high accuracy in Greek. Fine tuning of the Greek LLM in the domain of education can further improve the quality of responses in educational contexts and promises to enhance the user experience by providing more precise and relevant information tailored to the specific linguistic and educational needs of Greek-speaking users.

The third key area involves **improved prompt engineering to optimize AI efficiency and effectiveness**. Developing more precise and detailed prompts to guide AI responses, ensuring they are accurate and relevant to the users' queries, is essential. This feedback is consistent across all regions. Aligning lesson plans, quizzes, and other educational content with national or local curriculum standards to ensure relevance and compliance was highly suggested. This includes detailed grading rubrics, multiple correct answer options, adaptation to individual learning styles, and personalized feedback mechanisms. Additionally, ensuring adaptability to different educational levels, special educational needs and standards is highlighted. Implementing features that support collaborative learning and peer interaction, such as group projects, shared resources, and interactive discussions, was particularly highlighted by feedback from Sweden.

Co-funded by
the European Union

Another key area that emerged from the analysis of pilot results is the **enrichment of educational content and resources.** Users suggested the capability to integrate a wide range of their own educational resources, including external books, PDFs, and online materials. Incorporating multimodal features such as images, diagrams, videos, and audio aids was frequently mentioned across all regions as a way to make learning more engaging and interactive. Moreover, users propose the expansion of the range of subjects covered by SB to cater to a wider array of academic disciplines, e.g. tools specifically designed to solve math problems or tools for foreign language learning. However, these capabilities are currently limited by the project's original implementation methodology, which, being conceived in the pre-GPT era, primarily emphasized text processing technologies and methods.

Class management in TM and **practical utility** of both the SB and TM applications were two other key areas that collected significant recommendations. Integrating features that support the management of classes, including adding/removing students, tracking progress, and integrating with existing school systems, were noted. Compatibility with mobile devices, including smartphones, was also highlighted, addressing the needs of users who rely on mobile technology for educational activities, particularly noted by feedback from Sweden and Ireland.

Finally, the need for end users' **support and training** has emerged as a major finding of the pilot results. Providing detailed tutorials and user guides will help users navigate the various features and functionalities more effectively. Moreover, implementing training programs can significantly enhance user confidence and proficiency, supporting educators and students to leverage the capabilities of these tools. Building knowledge and skills in generative AI among teachers and students is expected not only to improve the acceptance of the tools but also to empower the end users to integrate them more efficiently into their teaching and learning processes. Enhanced support and training initiatives will ultimately lead to higher user satisfaction and more effective use of the educational technologies provided by SB and TM.

These findings will be aligned with the project's Description of Action, objectives, and scope, prioritized, and addressed as part of the WP5 activities, which aim to develop the final prototype of SB and TM, ensuring that the AI4EDU applications become more effective, user-friendly, and impactful educational tools.

Appendix

- Updated version of the SB testing and evaluation protocol: EN Student pilot protocol - manual.pdf.
- Updated version of the TM testing and evaluation protocol: EN Teacher pilot protocol - manual.pdf
- Students questionnaire in English: Evaluation of Conversational AI Assistant for Students (SB) (English version).pdf
- Teachers Questionnaire in English: Evaluation of Conversational AI Assistant for Teachers (TM) (English version).pdf
- Students Quantitative Data: Link, Students Qualitative Data: Link
- Teachers Quantitative Data: Link, Teachers Qualitative Data: Link

Co-funded by the European Union