



Evaluation of educational impact results

Date: 30 June 2025



**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 n°	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, Ton epikoinonion 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: D6.2

Version: 2.0

Authors: Theodoros Karafyllidis (UCY) & Marios Papaevripidou (UCY)

Dissemination status: PU - Public

Reviewers: Spyridoula Stamouli (ARC), Hamam Mokayed (LTU)

Project no. 101087451

AI4EDU

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	26/05/2025	-	Draft	Theodoros Karafyllidis
1.0	12/06/2025	ALL	Quantitative analysis completed	Theodoros Karafyllidis
1.1	19/06/2025	ALL	Additions	Theodoros Karafyllidis
1.2	26/06/2025	ALL	Review	Spyridoula Stamouli & Hamam Mokayed
2.0	27/06/2025	ALL	Final editing	Theodoros Karafyllidis

Validation			
No.	Action	Beneficiary	Date
1	Prepared	Theodoros Karafyllidis	19/06/2025
2	Approved	Spyridoula Stamouli & Hamam Mokayed	26/06/2025
3	Released	Theodoros Karafyllidis, Marios Papaevripidou, Spyridoula Stamouli	30/06/2025

All rights reserved.

The document is proprietary of the AI4EDU consortium members. No copying or distributing, in any form or by any means, is allowed without the prior written agreement of the owner of the property rights.

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

Executive Summary	6
Acronyms and abbreviations	7
1. Introduction.....	8
1.1. Project Background	8
1.2. Deliverable's Purpose	8
1.3. Relation to Other Work Packages and Deliverables	8
2. Overview of the Evaluation Framework.....	10
2.1. Evaluation Objectives and Scope	10
2.2. Research Questions.....	10
2.2.1. Student-related research questions	10
2.2.2. Teacher-related research questions:	11
2.3. Evaluation Design.....	11
2.4. Evaluation Instruments.....	12
3. Overview of pilot implementations.....	13
3.1. Pilot implementations in Cyprus.....	13
3.2. Pilot implementations in Greece	14
3.3. Pilot implementations in Ireland	14
3.4. Pilot implementations in Sweden.....	16
4. Data Analysis and Interpretation of Findings.....	20
4.1. Impact on Students' Academic Performance	20
4.1.1. Case Study 1: Physics Instruction (10th Grade, Cyprus)	20
4.1.2. Case Study 2: Physics Instruction (10th Grade, Cyprus)	22
4.1.3. Case Study 3: Chemistry Instruction (8th Grade, Cyprus)	23
4.1.4. Case Study 4: Biology Instruction (10th Grade, Greece)	24
4.1.5. Case Study 5: History Instruction (7th Grade, Greece).....	25
4.1.6. Case Study 6: Sustainability Citizenship (6th Grade, Ireland).....	26
4.1.7. Perceived Academic Impact: Student and Teacher Reflections	28

4.1.7.1.	Student Perspectives on Study Buddy's Academic Value	28
4.1.7.2.	Teachers' Reflections on Students' Academic Performance	33
4.2.	Motivation and Engagement	34
4.2.1.	Survey-Based Evaluation of Motivation and Engagement.....	34
4.2.1.1.	Within Group Analysis	35
4.2.1.2.	Between Group Analysis	37
4.2.1.3.	Between-Subject Analysis	38
4.2.1.4.	Demographic Analysis.....	40
4.2.2.	Student and Teacher Reflections on Motivation and Engagement	41
4.2.2.1.	Student Feedback on Motivation and Engagement.....	41
4.2.2.2.	Teacher Reflections on Students' Motivation and Engagement.....	43
4.2.3.	Concluding remarks	45
4.3.	Application Usage.....	46
4.3.1.	Study Buddy Log File Analysis	46
4.3.2.	Teacher Mate Log File Analysis	50
4.4.	Teacher Mate's Influence on Teacher Instructional Practice.....	53
4.4.1.	Teachers' Initial Practices, and Views on AI Use in the Classroom.....	53
4.4.2.	Teachers' Reflections on the Use of Teacher Mate	55
5.	Synthesis of Findings	57
5.1.	Student-Related Research Questions.....	57
5.2.	Teacher-Related Research Questions.....	60
6.	Conclusions and Recommendations	62
6.1.	Conclusions	62
6.2.	Recommendations for the Educational Community	63
6.3.	Policy Recommendations	64
6.3.1.	Final Remarks.....	65
7.	References	65

Executive Summary

This deliverable presents the results of the second pilot evaluation of the AI4EDU applications, Study Buddy (SB) and Teacher Mate (TM), conducted across primary and secondary school settings in Cyprus, Greece, Ireland, and Sweden. The evaluation aimed to assess the educational value of these AI-powered applications in authentic classroom environments, focusing on their impact on student learning, motivation, engagement, and teaching practices.

The findings are based on a comprehensive, mixed-methods approach involving academic tests, motivation and engagement surveys, usage log data, classroom observations, teacher interviews, and student feedback. This triangulated methodology enabled a comprehensive analysis of how the tools functioned in diverse educational contexts and under varying implementation conditions.

SB, the student-facing conversational AI assistant, showed particular promise in supporting content comprehension, personalized feedback, and learner autonomy, especially when used in teacher-facilitated, curriculum-aligned settings. Positive outcomes were most evident in inclusive classrooms and under conditions of structured integration, where students used SB for concept clarification, assessment preparation, and inquiry-based learning.

TM, the teacher-support platform, was found to be a useful resource for planning, assessment design, and differentiated instruction. Teachers who engaged with TM consistently reported time savings, inspiration for instructional content, and improved ability to meet diverse student needs. However, technical limitations and uneven training and engagement limited its broader uptake across countries.

Across both applications, the evaluation highlighted the critical role of teacher readiness, agency, and professional support in enabling successful AI integration. Classrooms where teachers had prior exposure to the applications, confidence in their pedagogical value, and the autonomy to adapt them to local contexts saw deeper and more sustained engagement. At the institutional level, organisational enablers such as distributed leadership, collaborative cultures, and psychological safety amplified the effectiveness of AI-supported practices.

Challenges also emerged, including underutilization of some features due to time and school schedule constraints and short pilot durations. Quantitative data from student self-reports proved inconclusive in isolation, underscoring the need for robust, triangulated approaches to evaluating educational technology.

The deliverable concludes with a set of strategic recommendations for the educational community and policy stakeholders. These stress the importance of sustained teacher training, inclusive tool design, school-wide readiness, and responsive policy frameworks. Findings from this evaluation will directly inform Work Package 7, which focuses on implementation guidelines and policy recommendations for equitable and pedagogically meaningful AI integration in education.

In sum, the pilot confirms that AI applications such as SB and TM can add significant educational value, especially when introduced within thoughtfully designed, teacher-led, and institutionally supported learning ecosystems.

Acronyms and abbreviations

Abbreviation	Description
AI	Artificial Intelligence
AI4EDU	Artificial Intelligence for Education (project short name)
MES	Motivation & Engagement Survey
SEN	Special Educational Needs
SB	Study Buddy
TM	Teacher Mate
WPs	Work Packages

1. Introduction

1.1. Project Background

The Conversational AI assistant for teaching and learning - AI4EDU project aims to enhance educational outcomes by integrating Artificial Intelligence (AI)-powered applications in real-world classroom settings. The project focuses on developing two key AI-driven applications: Study Buddy (SB), a conversational assistant for students, and Teacher Mate (TM), a platform designed to support teachers with instruction, assessment, and student monitoring. These applications are developed through iterative cycles of adaptation, testing, and refinement, driven by user requirements and feedback from both students and teachers.

The project also establishes a pedagogical framework that guides the design, implementation, and evaluation of AI applications for teaching and learning. Its methodology emphasizes ethical deployment, user-centered design, and continuous evaluation in diverse educational contexts across Europe. Pilot studies conducted in four countries (Cyprus, Greece, Ireland, and Sweden) assess both the usability and acceptance of these AI applications (WP4) and their impact on teaching and learning outcomes (WP6). The results will inform best practices, implementation guidelines, and policy recommendations to ensure equitable, transparent, and effective use of AI in education (WP7).

1.2. Deliverable's Purpose

This deliverable presents the final evaluation results of the AI4EDU applications, SB and TM, following their implementation in real-world educational contexts across four European countries: Cyprus, Greece, Ireland, and Sweden. It provides a comprehensive analysis of the applications' pedagogical impact, drawing on both qualitative and quantitative data collected during the second pilot cycle in WP6.

The document synthesizes findings related to student learning outcomes, motivation, engagement, and usage patterns of SB, as well as teachers' experiences with TM in supporting planning, assessment, and differentiated instruction. It also includes an in-depth exploration of contextual and organisational enablers that shaped the effectiveness of AI integration.

In addition to documenting the educational impact of the tools, this deliverable presents strategic recommendations for the educational community and policy stakeholders, aiming to support sustainable, inclusive, and pedagogically grounded AI integration in European classrooms. It concludes with key insights that inform future scaling efforts and provide a bridge to policy and implementation guidelines development in Work Package 7.

1.3. Relation to Other Work Packages and Deliverables

This deliverable builds upon and complements earlier outputs across multiple Work Packages (WPs) in the AI4EDU project, contributing to a cohesive, evidence-based progression from development to evaluation and policy formulation.

It is a direct continuation of Deliverable D6.1, which outlined the evaluation framework, instruments, and implementation plan for piloting SB and TM. The present deliverable operationalises and completes that framework by presenting empirical results and interpretive analysis based on classroom implementation across the four participating countries.

It draws on the pedagogical principles, user requirements, and ethical considerations outlined in WP2 (as detailed in D2.1 and D2.2), ensuring that the evaluation remains aligned

with the project's foundational vision. The iterative design and technical development efforts from WP3 and WP5 (D3.1, D3.2, D5.1, and D5.2) are also integral, as the applications were evaluated in their refined, classroom-ready forms.

Findings from this deliverable are further informed by the usability and technical evaluation activities carried out in WP4 (D4.2), ensuring that both pedagogical and technological aspects are considered in assessing real-world application and impact.

Crucially, the conclusions and recommendations presented here provide the empirical backbone for WP7, which focuses on the development of implementation guidelines and policy recommendations for AI in education. The insights from this deliverable will directly inform D7.1 ensuring that future outputs are rooted in robust classroom evidence and informed by stakeholder feedback gathered during pilot deployment.

In this way, the deliverable plays a central role in linking AI4EDU's research, development, and policy goals, supporting the project's broader mission to foster inclusive, equitable, and effective AI integration in European education systems.

2. Overview of the Evaluation Framework

This chapter summarizes the evaluation framework established in Deliverable D6.1 and explains how it informed the design and execution of the second pilot cycle of the AI4EDU applications. The framework was developed to assess the educational impact of SB and TM in real classroom settings across four European countries: Cyprus, Greece, Ireland, and Sweden. It provided a coherent structure for evaluating how these AI applications influence both teaching practices and student learning outcomes.

The evaluation design is grounded in experimental educational research and employs a mixed-methods approach, integrating both quantitative and qualitative data to provide a comprehensive, context-sensitive understanding of the AI4EDU applications' impact on teaching and learning. It is theoretically informed by three foundational learning theories (Constructivist Learning Theory, Self-Determination Theory, and Cognitive Load Theory) which guided both the development of the applications and the structure of the evaluation.

2.1. Evaluation Objectives and Scope

The overarching goal of the evaluation was to examine whether, and in what ways, the AI4EDU applications contribute to enhanced teaching and learning outcomes. The evaluation framework was designed to capture both student-centered and teacher-centered effects through a mixed-methods approach, integrating quantitative and qualitative data sources. Specific evaluation objectives included:

- Assessing the impact of SB on students' academic performance, motivation, and engagement.
- Evaluating the extent to which TM supports teachers in their lesson preparation, instructional delivery, and assessment practices.
- Identifying key conditions for effective implementation across different educational systems and national contexts.

These objectives reflect the project's broader ambition to validate the educational value of SB and TM and demonstrate their potential to improve the quality of teaching and learning.

2.2. Research Questions

The evaluation framework was guided by two complementary sets of research questions: one focusing on students as users of SB, and one focusing on teachers as users of TM. These questions reflect the dual focus of the AI4EDU applications on enhancing both learning and teaching processes.

2.2.1. Student-related research questions

1. How does SB support students in preparing for assessments and tracking their academic progress?
2. To what extent does SB improve students' understanding and retention of key concepts?
3. How does the use of SB impact student engagement and motivation in the learning process?
4. How do students perceive the usability and effectiveness of SB in enhancing their learning experience?
5. What learning patterns and trends can be observed from students' interactions with SB, and how do these correlate with their academic outcomes?

2.2.2. Teacher-related research questions:

1. How does TM affect teachers' efficiency in lesson planning, instructional delivery, and assessment tasks?
2. In what ways does TM support teachers in providing personalized instruction to address diverse student needs?
3. How effective is TM in helping teachers monitor student progress and identify areas for improvement?
4. What is the perceived impact of TM on reducing teacher workload?
5. How do teachers evaluate the overall impact of TM on student learning outcomes?

These questions provided a structured lens for evaluating the applications' educational value across different use cases and school contexts. They are based on the original framework presented in D6.1 *Methodology for the evaluation of educational impact*. However, it should be noted that some of these questions were slightly reformulated or interpreted more flexibly, as not all features of SB and TM were utilised uniformly across pilot sites. The variations in tool usage, influenced by local implementation conditions and teacher autonomy, required a degree of adaptation in the evaluation focus, while preserving the main aim of assessing the educational value of the applications. For example, due to time restrictions, features of SB and TM enabling longitudinal tracking of academic progress were not used, meaning that Student Q1 and Teacher Q3 were modified to focus on short-term use, e.g. drawing on the use of *Quiz generation* or *Self-assessment* tools.

2.3. Evaluation Design

The evaluation employed a mixed-methods experimental design, combining quantitative and qualitative data collection to assess the educational impact of the AI4EDU applications in real classroom settings.

For SB, a control–experimental group comparison approach was adopted:

- In the experimental group, students used SB either outside class (e.g., for self-study, homework, and test preparation) or in class (e.g., for activity completion, formative evaluation, project-based or feedback-based learning).
- The control group followed traditional study practices without the use of AI.

All participating teachers used TM to support their teaching in both the control and experimental groups, ensuring a consistent instructional baseline while isolating the impact of SB on student learning outcomes.

The evaluation was conducted across a range of subjects in secondary education¹, including science disciplines (Physics, Biology, Chemistry), humanities (History, Literature), and interdisciplinary areas such as Sustainability Citizenship. Although the evaluation adhered to a shared methodological framework, each participating country adapted its implementation to align with national curricula, institutional structures, and logistical considerations. These contextual adaptations enriched the evaluation process, offering a more nuanced and comprehensive understanding of the educational value of the AI4EDU applications.

The evaluation was organized into three key phases:

¹ In the case of Ireland, the evaluation also included 5th and 6th grade classes, which form the senior cycle of primary education. This stage was considered appropriate due to its strong alignment with early secondary education in terms of student developmental level and curricular content.

1. **Pre-implementation phase:** This preparatory phase focused on establishing the necessary conditions for effective and ethical implementation. Activities included the recruitment of teachers and students, informed consent collection, teacher and student training sessions, and administration of baseline instruments.
2. **Implementation phase:** During this phase, the AI4EDU applications were integrated into day-to-day teaching and learning practices. Teachers employed TM to support lesson planning, assessment, and instructional delivery, while students in the experimental groups engaged with SB either in class or at home, depending on the subject and national context.
3. **Post-implementation phase:** The final phase included the administration of post-tests and collection of qualitative feedback through surveys, interviews, and usage data.

2.4. Evaluation Instruments

A comprehensive set of evaluation instruments was developed to assess both the cognitive and affective dimensions of learning, as well as to capture user perceptions, application usage, and implementation fidelity. These instruments were designed to offer complementary insights into the educational impact of the AI4EDU applications across diverse teaching and learning environments. The toolkit included:

- **Academic pre- and post-tests:** Developed in collaboration with participating teachers to assess students' understanding of subject-specific content. These tests were tailored to align with national curricula and the learning objectives of each subject.
- **Motivation and Engagement Survey (MES):** Adapted from the validated framework of Lee & Reeve (2012), this 38-item Likert-scale survey explored psychological need satisfaction, self-efficacy, mastery goals, and multiple dimensions of engagement (behavioral, emotional, cognitive, and agentic).
- **Usage logs:** Automatically stored within the AI4EDU platform database, these logs captured user interactions, including chat frequency, duration, and activity type. This data provided valuable insights into how both teachers and students engaged with the applications in real educational contexts.
- **Student post-implementation questionnaire:** Designed to capture students' perceptions of SB, this questionnaire focused on its usefulness, ease of use, and role in supporting their learning experience.
- **Teachers' feedback:** Collected after implementation through interviews or questionnaires to explore teachers' reflections on the integration of TM, including its impact on instructional practices, workload, and perceived student outcomes.
- **Classroom observation protocols:** Structured templates and frameworks, such as SWOT (Strengths, Weaknesses, Opportunities, Threats) and NOISE (Needs, Opportunities, Improvements, Strengths, Exceptions), were used to document classroom dynamics, teacher-student interactions, and the fidelity of AI integration into instructional practices.

Although the full set of evaluation instruments was made available to all participating teams, their actual use varied across countries and cases due to contextual factors such as class size and teacher availability. This flexible implementation strategy allowed each pilot to adapt to its local educational environment while still aligning with the overarching evaluation framework. It also helped minimize disruptions to schools' daily operations and ensured respect for existing planning and pedagogical routines.

3. Overview of pilot implementations

This chapter provides an overview of how the second cycle of pilot implementations of the AI4EDU applications was carried out across the four participating countries: Cyprus, Greece, Ireland, and Sweden. Guided by a common evaluation framework, the piloting activities were shaped by shared objectives, instruments, and procedures, while also reflecting the specific institutional, curricular, and organizational conditions of each national context.

The following sections present descriptive accounts of the national implementations, including teacher recruitment, training procedures, classroom integration strategies, and contextual factors that influenced the execution of the evaluation activities. These summaries offer insight into the practical realities of conducting educational interventions across varied school systems.

3.1. Pilot implementations in Cyprus

In Cyprus, eight secondary school teachers responded to an open call for participation in the second cycle of the AI4EDU pilot and attended an online training session held on 31 January 2025. The session introduced participants to the project's objectives and requirements, provided an overview of TM and SB, and familiarized them with the planned evaluation procedures and instruments. The group included three Physics teachers and five teachers from various disciplines, including Biology, Chemistry, Mathematics, English Literature, and Greek Literature. All teachers were affiliated with public secondary education schools.

Out of the eight trained teachers, four proceeded with classroom implementations involving their students. These included two Physics teachers, one Chemistry teacher, and one English Literature teacher. Each teacher selected subject-relevant content for the pilot and used TM to design corresponding lesson materials, develop evaluation tests, and set up virtual classrooms. They then divided their students into experimental and control groups, aiming for relatively balanced group sizes. An overview of the participating teachers, subjects, grade levels, and student group sizes is presented in Table 3.1.

After assigning students to groups, teachers distributed information sheets and consent forms required for participation in the study. This step caused significant delays, as many students either failed to return the forms or submitted them late. Once consents were obtained, teachers administered the content-specific pre-tests and the Motivation and Engagement Survey. Students in the experimental groups were then given access credentials for SB and received guidance on how to use the application.

Pilot implementations began in mid to late February 2025. However, the timeline was affected by national holidays and celebratory events in March, which interrupted regular instruction and extended the activities until mid-May. This period also overlapped with exam preparation, further complicating efforts to maintain consistent implementation.

While the evaluation framework was built on a shared methodological structure, the Cypriot implementations featured context-specific adaptations. Most teachers introduced SB as a self-study application for homework and exam preparation. However, the English Literature teacher integrated the application into classroom instruction and developed custom activities to align with her teaching goals.

The administration of evaluation instruments varied across cases. In some classes (e.g., Chemistry and one Physics group), MES was administered in paper format, while others (e.g., English and the second Physics group) utilized an online version. Academic performance

tests were administered in all participating classes; in the case of English Literature, the test for the experimental group was delivered through the AI4EDU platform.

Following the implementation period, usage logs from SB and TM were collected, along with student responses to the Post-Implementation Questionnaire, to assess engagement and user experience. Similarly, teachers provided feedback reflecting on their use of TM.

3.2. Pilot implementations in Greece

In Greece, eleven secondary school teachers participated in the training event for the second cycle of the AI4EDU pilot. The training was conducted in person on 28 January 2025 at Ellinogermaniki Agogi (EA) private school in Athens. The session provided an overview of the project's objectives, introduced the AI4EDU applications, and familiarized participants with the planned evaluation procedures. Since EA is a project partner, many of the teachers were already familiar with SB and TM and had participated in the first phase of pilots, which evaluated their usability and technology acceptance. The participating group included nine male and two female teachers from diverse subject areas, including Biology, History, Physics, Chemistry, and Literature, spanning both lower and upper secondary education levels.

Out of the eleven trained teachers, two proceeded with implementing the pilot activities in their classrooms: one Biology teacher and one History teacher. The Biology teacher conducted the pilot with four classes of first-year upper secondary students, forming two experimental and two control groups. The History teacher implemented the pilot with three classes of first year lower secondary students, organizing them into two experimental groups and one control group. A summary of the participating groups is included in Table 3.1.

The implementations began in the second week of February 2025 and concluded mid-May. As in Cyprus, the timeline was influenced by national holidays and the Easter break, which temporarily interrupted regular instruction. Nevertheless, the teachers were able to maintain the core structure of the pilot activities.

In the Biology case, the teacher introduced content and key concepts during classroom instruction. Students in experimental groups then used SB at home to further explore the material, respond to assigned questions, and clarify unfamiliar terms. In follow-up lessons, students discussed their findings with peers and teachers, allowing them to consolidate their understanding through a combination of individual study and guided discussion.

In contrast, the History implementation was carried out primarily in the school's computer lab. Students used SB to search for information related to Byzantine emperors, with a focus on understanding their strategies and governance. Teachers emphasized the importance of critically evaluating the platform's responses and cross-referencing them with textbooks and other authoritative sources. Students presented their findings in class using the Canvas platform, fostering historical thinking and digital literacy.

The administration of evaluation instruments varied slightly between the two cases. In History, students completed the MES online (both pre- and post-implementation), a paper-based academic performance test, and the student post-implementation questionnaire. In Biology, a content specific post-test was administered on paper and students also completed the post-implementation questionnaire. In both cases, usage logs from SB and TM were collected to provide insight into user interaction patterns.

3.3. Pilot implementations in Ireland

In Ireland, the AI4EDU pilots were implemented at St. Thomas' Senior National School, a DEIS-designated (Delivering Equality of Opportunity in Schools) institution known for its

inclusive ethos and strong commitment to sustainability and digital learning initiatives. The school's active involvement in the Green Schools initiative and Education for Sustainable Development (ESD) programmes, combined with its technological infrastructure and pedagogical innovation, positioned it as an ideal environment for piloting the AI4EDU applications within diverse learning environments. Importantly, the school is not considered as a conventional pilot site but has integrated the TM and SB applications into its regular teaching practice, serving as a sustained evaluation environment for their ongoing use and development within authentic classroom settings.

The main selected curriculum area was Sustainability Citizenship, aligning with national priorities for Social, Environmental, and Scientific Education (SESE) in Irish primary and post-primary education. SESE is an interdisciplinary domain encompassing Science, Geography, and History, and supports cross-curricular themes, such as environmental awareness, active citizenship, and climate literacy. Selected curriculum areas also align with the Education for Sustainable Development (ESD), an approach that equips students with the knowledge, skills, and values needed to contribute to a sustainable future. Topic selection was guided by the Irish national strategy on ESD, UNESCO's Sustainable Development Goals (SDGs), and national initiatives such as the Green Schools programme and the Climate Action framework.

The pilots were implemented in two phases. The first phase took place from 20 to 22 January 2025, involving two 6th-grade classes (28 boys and 20 girls, totaling 48 students) and six teachers. Each class was divided into control and experimental groups, with three teachers assigned to each. In the control group, teachers employed traditional instructional methods, including PowerPoint presentations, worksheets, and guided discussions. In contrast, the experimental group integrated the AI4EDU applications into classroom activities. Teachers used TM to plan lessons, generate content, and design activities, while students used SB to explore, summarise, and personalise learning content and complete tasks such as quizzes. The AI-supported activities encouraged deeper reflection and allowed students to explore their own contributions to sustainability goals within their school environment.

Teachers and several students were already familiar with the AI4EDU applications from earlier participation, so no additional training was necessary for them. This prior experience facilitated a streamlined and effective implementation. Non-trained students received initial training from their teachers, as part of regular classroom activities. All teaching occurred within regular class hours, and SB was used exclusively in the classroom, with no homework-based activities included.

The implementation followed a three-phase format: a pre-test (5 minutes), an instructional period (35 minutes), and a post-test (5 minutes).

Before and after the implementation, students in both the control and experimental groups completed an online test designed to assess their understanding of key concepts related to Green Schools, including sustainability, climate literacy, and foundational scientific principles. The pre-test specifically aimed to establish baseline knowledge and attitudes toward Green School principles and Sustainability Citizenship. A total of 21 students from the control group and 20 from the experimental group completed the pre-test, while 23 control group students and 19 experimental group students completed the post-test.

Mixed-methods evaluation complemented this quantitative data, using SCOT and NOISE frameworks. Observations were co-conducted by teachers and Drumcondra Education Centre (DEC) staff and were coded against UNESCO's AI Competency Framework and ethical indicators from the EU AI Act.

The second pilot implementation phase took place on 8–9 May 2025. While maintaining the same general methodology, this implementation expanded in scope to encompass a broader range of students, staff roles, and curricular applications. It involved one 5th-grade and two 6th-grade classes, comprising approximately 72 students (39 boys and 33 girls), three classroom teachers, and additional school personnel, including an English as an Additional Language (EAL) specialist, a Special Educational Needs (SEN) teacher, a senior member of the management team, and two representatives from the DEC, members of the AI4EDU team. This enabled a richer evaluation of the AI4EDU applications, particularly in relation to their impact on diverse learner populations, including Ukrainian refugee students, Roma and Traveler students, and those with SEN.

The curricular scope was also expanded in this phase. In the area of Humanities, TM and SB were used to support narrative development, vocabulary acquisition, and debating skills. Within the SESE domain, SB enabled students to engage in ecological thinking and helped them identify their roles in promoting sustainability. In English Language instruction, both applications were employed to support structured argumentation and creative writing, including poetry across various genres.

Unlike the first phase, no pre- or post-tests were administered during this phase to measure academic outcomes. Instead, the focus was placed on broader pedagogical and organizational insights. Observations and participant feedback were analysed using the SCOT and NOISE frameworks. Additional survey instruments were introduced to assess aspects of school culture, leadership dynamics, and team collaboration. The evaluation framework drew on key constructs from organizational psychology, such as Edmondson's concept of Psychological Safety and Self-Determination Theory, and aligned its findings with UNESCO's AI Competency Framework and the EU AI Act's ethical guidelines for AI use in education.

All data collection activities in both phases adhered strictly to the General Data Protection Regulation (GDPR), the school's Acceptable Use Policy (AUP), and the provisions of the EU AI Act. Teachers retained autonomy over lesson pacing and the integration of AI applications, allowing flexible adaptation to classroom dynamics. No significant technological issues were reported; minor technical challenges were addressed promptly through on-site support.

It is worth noting that although the core implementation activities were concentrated over three days in January and two days in May, the use of TM and SB continued beyond these formal sessions. This ongoing use generated valuable log data, offering further insight into how the applications were integrated into everyday teaching and learning. The combined data from both pilot phases provided a comprehensive foundation for evaluating the sustained impact and classroom relevance of the AI4EDU applications.

3.4. Pilot implementations in Sweden

In Sweden, the AI4EDU pilot was conducted in collaboration with Björknäsgymnasiet, a secondary school located in Boden. The implementation focused on two subject areas: Swedish Literature and Physics. The Swedish Literature pilot involved one teacher and a total of 30 students, divided equally into experimental and control groups. In Physics, the pilot began with one teacher and ten students (five per group), but the implementation was discontinued due to the teacher's prolonged illness.

The pilot was structured across three main phases:

Pre-Implementation Phase: Preparatory activities began in early January 2025. On January 1st, LTU shared consent forms with the school's point of contact, and by January 10th, Björknäsgymnasiet had identified the participating teachers and shared the relevant

teaching materials, such as chapters, academic papers, and exercises in PDF format, to be covered during the pilot period. By January 17th, LTU provided participants with user guide videos introducing the AI4EDU applications, SB and TM, to support both teachers and students in navigating the platforms.

Implementation Phase: The implementation formally began with two half-day onboarding sessions: on January 22nd, LTU held an interactive training session with approximately 24 students, followed by a session on January 23rd for seven interested teachers from different subject areas. Classroom implementations at Björknäsgymnasiet commenced on January 27th and continued until February 21st, 2025. During this period, the Swedish Literature teacher facilitated both in-class and at-home student engagement with SB, using the platform to deepen understanding of key literary themes and support critical analysis of selected texts. Students in the experimental group were assigned specific activities via SB, while the control group followed the standard instructional approach.

Post-Implementation Phase: From February 24th to 28th, LTU administered post-questionnaires to gather data on student experiences. A follow-up meeting with participating teachers was held to reflect on the pilot's outcomes and gather qualitative insights. Final data analysis and reporting were completed by mid-May, incorporating usage logs from both SB and TM, student questionnaires, and feedback from educators.

While the Physics implementation could not be completed, the Swedish Literature pilot provided valuable insights into the integration of AI-based tools in upper secondary classrooms in Sweden. The experience also highlighted the importance of adaptability and teacher engagement in ensuring the success of such pilots.

Table 3.1. Overview of Participating Teachers and Student Groups

No.	Country	Subject(s)	Grade Level	Topic	Experimental Group Size	Control Group size
1	Cyprus	Physics	10 th grade	Work/ Kinetic Energy	16	20
2	Cyprus	Physics	10 th grade	Gravitational Force	23	24
3	Cyprus	English literature	11 th grade	Text production	20	20
4	Cyprus	Chemistry	8 th grade	Chemical elements/ compound	21	20
5	Greece	Biology	10 th grade	Nervous System	47	41
6	Greece	History	7 th grade	Byzantine Emperors	45	25
7	Ireland	Sustainability Citizenship	6 th grade	Green Schools	24	24
8	Ireland	English Language, Humanities, SESE topics	5 th & 6 th grade	Narrative, Poetry, Debate, Sustainability	36	36
9	Sweden	Physics	10 th grade	Electrophysics	5	5
10	Sweden	Swedish Literature	10 th grade	Enlightenment	15	15



Image 3.1: Pilot Implementation in Greece



Image 3.2: Pilot Implementation in Ireland

4. Data Analysis and Interpretation of Findings

This chapter presents the results of the second cycle of pilot implementations of the AI4EDU applications, SB and TM. The analysis focuses on the key impact areas defined in the evaluation framework, namely student academic performance, motivation and engagement, application usage, and teaching and learning practices. By triangulating data from multiple sources, including academic pre- and post-tests, the Motivation and Engagement Survey, post-implementation questionnaires, application usage logs, classroom observations, and teacher feedback, the chapter offers a comprehensive view of the educational impact of the AI4EDU applications.

To ensure coherence and facilitate comparison, findings are organized thematically rather than by country. Each section focuses on a specific impact domain and integrates quantitative data with qualitative insights where available. While some variation exists across national implementations in terms of data availability and context, the analysis emphasizes common trends, contextual adaptations, and emergent themes, contributing to a deeper understanding of how AI technologies function in real-world educational environments.

4.1. Impact on Students' Academic Performance

This section examines the impact of the AI4EDU applications on students' academic performance, based primarily on data collected through pre- and post-implementation academic tests. These quantitative results are complemented by qualitative insights drawn from student post-implementation questionnaires, teacher feedback, and classroom observation notes. The analysis is organized into focused mini case studies, each reflecting a distinct disciplinary and national context. In addition to test-based outcomes, the section incorporates students' and teachers' perspectives on perceived learning gains, offering a more comprehensive understanding of the educational effects of the applications.

4.1.1. Case Study 1: Physics Instruction (10th Grade, Cyprus)

In Cyprus, a Physics teacher participated in the second cycle of the AI4EDU pilot with two 10th-grade classes. The unit focused on Work and Kinetic Energy. The teacher organized his students into two groups: an experimental group consisting of 16 students and a control group of 20 students. Group assignment followed natural classroom division, with both groups being comparable in size and academic performance.

The intervention lasted approximately three weeks. During this period, students in the experimental group used SB as a **self-study tool outside of class**, primarily to review concepts and prepare for assessments. In contrast, students in the control group relied on traditional resources, such as textbooks and online materials, without access to SB. At the same time, the teacher used TM to support lesson planning and generate curriculum-aligned content and test questions.

To evaluate the educational impact of SB, both groups completed the same academic performance test at the end of the unit. The students' mean test scores are summarized in Table 4.1.

Table 4.1: Descriptive Statistics of Post-Test Scores by Group (Case Study 1)

Group	Group size	Mean score /10	Standard deviation
Control	20	6,05	1,85
Experimental	16	6,94	1,84

As shown in Figure 4.1, students in the experimental group had a higher mean post-test score than those in the control group, suggesting a potential learning benefit for students who used SB.

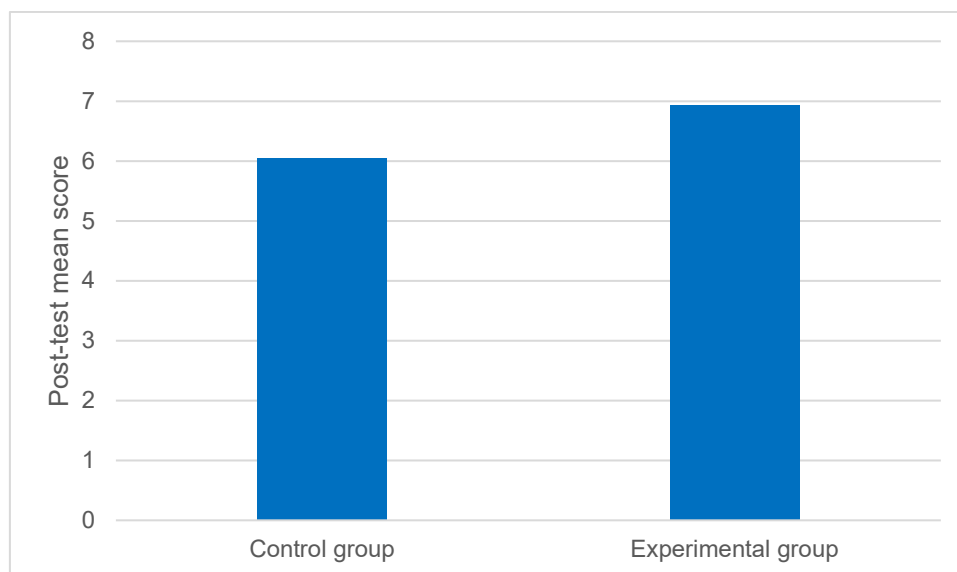


Figure 4.1: Distribution of Post-Test Scores by Group (Case Study 1)

Before selecting the appropriate test to assess the statistical significance of this difference, a Shapiro-Wilk test was performed to check for normality. As shown in Table 4.2, both groups' score distributions significantly deviated from normality ($p < 0,05$), warranting the use of non-parametric methods for comparison.

Table 4.2: Shapiro-Wilk Test for Normality (Case Study 1)

Group	Statistic	Degrees of freedom (df)	Significance (p)
Control	0,898	20	0,037
Experimental	0,847	16	0,012

Since both distributions significantly deviated from normality ($p < 0,05$), a Mann-Whitney U test was performed (Table 4.3).

Table 4.3: Mann-Whitney U Test Results (Case Study 1)

Test	Mann-Whitney U value	Significance (p)
Mann-Whitney U test	112,5	0,132

Although the experimental group outperformed the control group by nearly one point on average, the difference was not statistically significant, suggesting that this result cannot be confidently attributed to the intervention.

Nevertheless, the higher mean score of the experimental group points to a positive trend in favor of using SB as a self-study tool, highlighting its potential to support student learning when integrated thoughtfully into out-of-class study routines.

4.1.2. Case Study 2: Physics Instruction (10th Grade, Cyprus)

A second Physics teacher in Cyprus also participated in the AI4EDU pilot with two 10th-grade classes, focusing on Gravitational Force. Students were divided into two groups: a control group of 24 students and an experimental group of 23 students. The division followed existing classroom structures and reflected comparable group sizes and prior academic achievement.

The implementation extended over approximately three weeks. During this period, students in the experimental group were given access to SB, which they used independently **outside of school** to review the topic and consolidate their understanding. In contrast, students in the control group relied exclusively on conventional study methods, including textbooks, notes, and web resources, without using SB. Meanwhile, the teacher used TM to support the design of lesson content, generate test items, and manage instructional materials aligned with the curriculum.

To assess learning outcomes, both the control and experimental groups completed a content-specific academic test before and after the instructional unit. The difference between post-test and pre-test mean scores was used to calculate each group's mean gain score, which served as the primary indicator of academic progress. Descriptive statistics for the gain scores are presented in Table 4.4.

Table 4.4: Descriptive Statistics of Pre-Test, Post-Test, and Gain Scores by Group (Case Study 2)

Group	Group size	Pre-test mean score /100	Post-test mean score /100	Mean gain score /100
Control	24	37,00	59,17	22,17
Experimental	23	40,87	71,83	30,96

As shown, the experimental group achieved a notably higher mean gain score compared to the control group, suggesting that students who used SB may have benefited more from the instructional period.

To determine statistical significance of this difference, the Shapiro-Wilk test was first conducted to assess the normality of the data. As shown in Table 4.5, in both cases the gain scores were normally distributed ($p > 0,05$), allowing for the use of parametric tests.

Table 4.5: Shapiro-Wilk Test of Normality (Case Study 2)

Group	Statistic	Degrees of freedom (df)	Significance (p)
Control	0,979	24	0,875
Experimental	0,935	23	0,140

According to the independent samples t-test (Table 4.6), the results are statistically significant at the 0,05 level, indicating that students in the experimental group achieved significantly higher learning gains compared to their peers in the control group. This supports the effectiveness of SB in enhancing student understanding when integrated as a supplementary learning tool. The engagement patterns reflected in the students' usage data provide a meaningful interpretation for the demonstrated learning gains (see Section 4.3.1).

Table 4.6: Independent Samples t-Test Results (Case Study 2)

Test	t value	Significance (p)
Independent Samples t-test	-3,56	<0,001

4.1.3. Case Study 3: Chemistry Instruction (8th Grade, Cyprus)

In Cyprus, a Chemistry teacher implemented the AI4EDU applications with two 8th-grade classes during the second pilot cycle. The instructional unit focused on the topics of chemical elements and chemical compounds. The teacher divided her students into two groups according to their regular class structure: an experimental group consisting of 21 students and a control group of 20 students. Both groups were similar in size and prior academic level.

The implementation spanned approximately one month. During this period, students in the experimental group used the SB application independently **outside the classroom** to reinforce their understanding of the taught content. Meanwhile, the control group followed a traditional learning path, utilizing standard classroom materials and resources without the AI4EDU applications. The teacher also made use of TM to prepare and organize her teaching materials and generate test items aligned with the curriculum.

To evaluate the impact of SB on students' academic performance, the teacher administered the same content-specific academic test to both groups before and after the instructional period. The mean gain score, calculated as the difference between the post- and pre-test results for each group, served as the primary indicator of learning progress. Descriptive statistics are presented in Table 4.7.

Table 4.7: Descriptive Statistics of Pre-Test, Post-Test, and Gain Scores by Group (Case Study 3)

Group	Group size	Pre-test mean score /20	Post-test mean score /20	Mean gain score /20
Control	18	5,50	11,22	5,72
Experimental	18	5,72	12,72	7,00

As shown in the table above, both groups demonstrated notable learning gains over the course of the instructional period. The experimental group, however, achieved a higher mean gain score, suggesting a possible advantage associated with the use of SB.

To determine whether this difference was statistically significant, a Shapiro-Wilk test was first conducted to assess the normality of gain scores in each group. The results (Table 4.8), showed no significant deviation from normality in either group, allowing the use of a parametric test. Subsequently, an independent samples t-test was conducted to compare the gain scores. As shown in Table 4.9, the difference between the groups was not statistically significant at the 0,05 level.

Table 4.8: Shapiro-Wilk Test of Normality (Case Study 3)

Group	Statistic	Degrees of freedom (df)	Significance (p)
Control	0,936	18	0,246
Experimental	0,983	18	0,978

Table 4.9: Independent Samples t-Test Results (Case Study 3)

Test	t value	Significance (p)
Independent Samples t-test	-0,979	0,335

Despite the observed numerical difference in average learning gains, the results do not reach statistical significance. This suggests that while the use of SB may have contributed to improved performance, the effect was not strong enough in this context to conclusively demonstrate a significant educational impact. Nevertheless, the trend remains favorable and aligns with the positive patterns seen in other case studies.

4.1.4. Case Study 4: Biology Instruction (10th Grade, Greece)

In Greece, one 10th-grade Biology teacher participated in the second cycle of the AI4EDU pilot, implementing the intervention across four separate classes. The instructional unit focused on the human nervous system. The teacher organized the classes into two experimental groups (25 students each) and two control groups (with 23 and 22 students, respectively). Group allocation followed the school's existing classroom structure, ensuring comparable group sizes and similar prior academic performance levels.

The implementation lasted approximately three weeks. During this period, students in the experimental groups used SB at home as a self-regulated learning tool, supporting their understanding of the topic through guided exploration and clarification of key biological concepts. Students in the control groups engaged with the same curricular content using traditional study approaches, without access to SB. Meanwhile, the teacher used TM to prepare lesson content, design classroom activities, and generate assessment materials aligned with the official curriculum.

To assess student learning outcomes, the same content-specific post-test was administered to all four classes at the end of the implementation period. The results are summarized in Table 4.10.

Table 4.10: Descriptive Statistics of Post-Test Scores by Group (Case Study 4)

Group	Group size	Mean score /20	Standard deviation
Control groups	41	13,02	4,79
Experimental groups	47	14,15	3,71

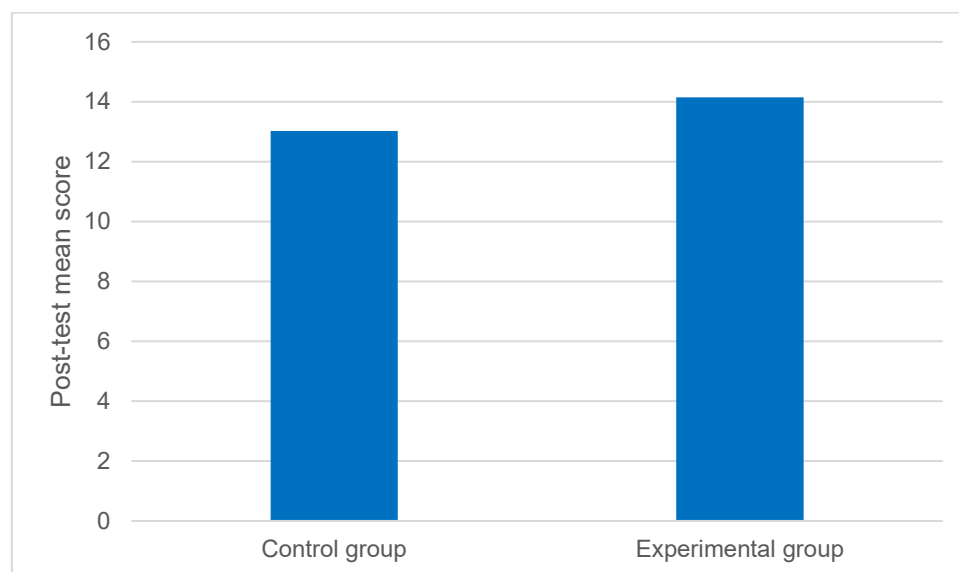


Figure 4.2: Distribution of Post-Test Scores by Group (Case Study 4)

As shown in Figure 4.2, students in the experimental groups achieved a higher mean score than those in the control groups, indicating a potential learning benefit associated with the use of SB.

To assess whether this observed difference was statistically significant, a Shapiro-Wilk test was first conducted to evaluate the normality of score distributions in both groups. The results are presented in Table 4.11.

Table 4.11: Shapiro-Wilk Test of Normality (Case Study 4)

Group	Statistic	Degrees of freedom (df)	Significance (p)
Control	0,913	41	0,004
Experimental	0,958	47	0,090

The results show that the distribution of scores in the control group significantly deviated from normality ($p < 0,05$), while the distribution in the experimental group was marginally normal. Given this, a non-parametric Mann-Whitney U test was deemed more appropriate for comparison (Table 4.12).

Table 4.12: Mann-Whitney U Test Results (Case Study 4)

Test	Mann-Whitney U value	Significance (p)
Mann-Whitney U test	1067	0,386

Although the experimental group demonstrated a higher average score, the difference between the groups was not statistically significant at the 0,05 level. This suggests that, in this case, the observed improvement may be attributable to chance rather than a definitive effect of the intervention.

Nonetheless, the trend remains positive and consistent with results observed in other case studies, reinforcing the potential of AI-supported applications like SB to support student learning in Biology when used as a supplementary resource.

4.1.5. Case Study 5: History Instruction (7th Grade, Greece)

In Greece, a 7th-grade History teacher participated in the second cycle of the AI4EDU pilot, implementing the intervention across three separate classes. The instructional unit focused on Byzantine emperors. Students were divided into one control group (25 students) and two experimental groups (25 and 26 students, respectively). The group assignment followed the school's existing classroom structure, with similar group sizes and relatively balanced prior academic performance.

The implementation lasted approximately three weeks. During this period, students in the experimental groups used SB in class to search for information about the Byzantine emperors. To foster historical thinking skills, the teacher encouraged students to evaluate the reliability of the AI-generated responses and to cross-reference the information with their textbooks and other trusted sources. Students in the control group engaged with the same content using only traditional resources, such as textbooks and printed worksheets, without access to the AI applications.

To evaluate learning outcomes, all students completed the same content-specific post-test at the end of the instructional period. Descriptive statistics are presented in Table 4.13.

Table 4.13: Descriptive Statistics of Post-Test Scores by Group (Case Study 5)

Group	Group size	Mean score /20	Standard deviation
Control groups	25	14,28	2,76
Experimental groups	45	17,09	2,71

As shown in Figure 4.3, students in the experimental groups achieved notably higher scores on average than those in the control group.

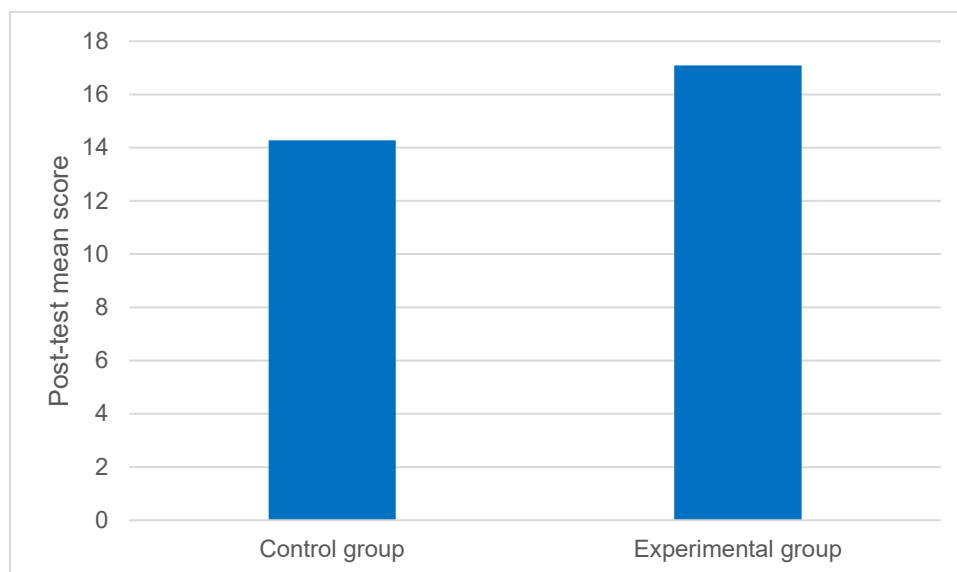


Figure 4.3: Distribution of Post-Test Scores by Group (Case Study 5)

To assess whether the observed differences were statistically significant, a Shapiro-Wilk test was first conducted to examine the normality of score distributions. The results are shown in Table 4.14. Both distributions deviated significantly from normality ($p < 0,05$). Accordingly, a non-parametric Mann-Whitney U test was conducted to compare post-test scores between the two groups (Table 4.15).

Table 4.14: Shapiro-Wilk Test of Normality (Case Study 5)

Group	Statistic	Degrees of freedom (df)	Significance (p)
Control	0,905	25	0,023
Experimental	0,878	45	<0,001

Table 4.15: Mann-Whitney U Test Results (Case Study 5)

Test	Mann-Whitney U value	Significance (p)
Mann-Whitney U test	256	<0,001

The test results revealed a statistically significant difference in post-test performance between the experimental and control groups, with students who used SB outperforming their peers. These findings indicate that the integration of AI tools, particularly when combined with critical thinking strategies and structured instructional support by their teacher in class, can significantly enhance students' conceptual understanding in the domain of History. In addition, usage data shed light on users' interaction patterns with the tool, revealing meaningful engagement behaviors that likely contributed to the observed learning gains (see section 4.3.1.).

4.1.6. Case Study 6: Sustainability Citizenship (6th Grade, Ireland)

In Ireland, the second pilot cycle involved six teachers and two 6th-grade classes, implemented within the interdisciplinary subject area of Sustainability Citizenship. The instructional unit focused on the theme of Green Schools. Each class was divided into an experimental group (24 students) and a control group (24 students), with three teachers supporting each group.

The implementation took place from 20 to 22 January 2025. During this period, teachers in the control group employed conventional teaching methods, including PowerPoint presentations, printed worksheets, and classroom discussions. In contrast, the experimental group integrated the AI4EDU applications into their instruction. Teachers used TM to design differentiated lesson materials, while students used SB to explore, understand, summarise, and personalise learning content. The lesson followed a structured format comprising a pre-test, a 35-minute instructional period, and a post-test.

More specifically, the instructional period centered on a lesson titled “School as a Green School and a Learning Ecology Promoting Sustainability Citizenship.” The session began with guiding questions, such as “What makes a school ‘green?’” and “Why focus on sustainability?”, used by teachers in both control and experimental groups to activate prior knowledge. This was followed by a teacher-led presentation of key concepts. In the control group, the teacher used a PowerPoint presentation and the whiteboard to introduce Green School themes, focusing on areas such as waste management, energy conservation, and the importance of green spaces. In the experimental group, the same content was delivered using TM tools, which supported term simplification, concept explanation, key point extraction, and the creation of differentiated materials tailored to students with special educational needs and those with English as an additional language.

Next, students participated in an interactive activity designed to help them reflect on their individual roles in promoting sustainability. In the control group, students completed printed worksheets asking them to identify sustainable practices in their daily lives. They then discussed their responses in small groups, with selected answers shared and noted on the classroom whiteboard. In the experimental group, TM was used to create individualised quizzes, group projects, and reflection activities. Students engaged with SB, receiving real-time feedback and suggestions to improve and clarify their responses.

The instructional period concluded with a reflection task focused on applying the concept of Learning Ecology to students’ own behaviours. In the control group, the teacher guided a discussion where students worked in pairs or small groups to choose actions, such as recycling or conserving water, that could contribute to the school’s ecological well-being. Meanwhile, in the experimental group, TM directed the reflection through tailored prompts, encouraging students to identify specific ways they could support the school’s Learning Ecology. SB helped students come up with ideas, suggesting simple actions they could take, and provided revisions to their written suggestions.

All instruction was delivered during regular class hours, and the applications were used exclusively in classroom settings, under the supervision and assistance of the teacher.

To assess the impact of SB on students’ academic performance, both groups completed the same content-specific test before and after the instructional period. The difference between pre- and post-test mean scores was calculated as the mean gain score, which served as the primary indicator of learning progress. Descriptive statistics for both groups are presented in Table 4.16.

Table 4.16: Descriptive Statistics of Pre-Test, Post-Test, and Gain Scores by Group (Case Study 6)

Group	Group size	Pre-test mean score /10	Post-test mean score /10	Mean gain score /10
Control	21	8,36	8,30	-0,06
Experimental	19	8,33	8,36	0,03

The results indicate that both groups performed at a high level in the pre-test, suggesting a strong initial understanding of the topic. Post-test scores remained similarly high, resulting in minimal changes in the mean gain scores. The experimental group showed a marginal positive gain (+0,03), while the control group experienced a slight decline (-0,06).

To determine whether this difference was statistically significant, a Shapiro-Wilk test was first conducted to assess the normality of gain scores in each group. The results (Table 4.17), showed no significant deviation from normality in either group, allowing the use of a parametric test. Subsequently, an independent samples t-test was conducted to compare the gain scores. As shown in Table 4.18, the difference between the groups was not statistically significant at the 0,05 level.

Table 4.17: Shapiro-Wilk Test of Normality (Case Study 6)

Group	Statistic	Degrees of freedom (df)	Significance (p)
Control	0,958	21	0,481
Experimental	0,952	19	0,405

Table 4.18: Independent Samples t-Test Results (Case Study 6)

Test	t value	Significance (p)
Independent Samples t-test	-0,077	0,939

These findings suggest that the brief intervention did not lead to measurable differences in academic performance between the groups. Given the high pre-test scores and the short duration of the implementation, significant learning gains were not expected. However, the slight improvement in the experimental group, although not statistically significant, may reflect the potential of AI-supported applications like SB to maintain or reinforce content understanding, particularly when integrated within a well-structured and differentiated instructional framework. Moreover, qualitative data gathered through post-implementation interviews with students and teachers as well as rich usage data (section 4.3.1.) offer valuable insights into perceived differences in academic performance, engagement, and motivation (sections 4.1.7.1.3. and 4.1.7.2.) as well as evidence of how patterns of use correlate with more sustained engagement and pedagogically meaningful application of SB tools.

4.1.7. Perceived Academic Impact: Student and Teacher Reflections

To complement the objective test score findings presented in the previous case studies, this section synthesizes qualitative feedback from both students and teachers regarding the perceived academic impact of the AI4EDU applications. The analysis draws on responses to post-implementation student questionnaires, interviews, and teacher reflections collected following the pilot implementations.

4.1.7.1. Student Perspectives on Study Buddy's Academic Value

In addition to test-based performance data, students' perceptions of SB's academic usefulness were captured through 161 post-implementation questionnaire responses from Cyprus (66), Greece (80), and Sweden (15), along with interview data gathered from experimental group students in Ireland.

4.1.7.1.1. Student Feedback from Closed-Ended Survey Responses (Cyprus, Greece, Sweden)

Three Likert-scale items in the questionnaire explored students' perceptions of SB's support in relation to (a) understanding the subject material, (b) organization and time management, and (c) academic confidence.

- **Understanding of Subject Material**

One key item (Q4a) asked students to indicate the extent to which they felt SB helped them better understand the subject material. Of the 161 respondents, 53 students (33%) agreed or strongly agreed, 51 (32%) remained neutral, and 57 (35%) disagreed or strongly disagreed (Figure 4.4). These results indicate a mixed reception of SB's academic value. While a notable portion of students perceived clear benefits, others were uncertain or unpersuaded. This mixed reception likely reflects the varied implementation conditions across countries. In most of the cases in Cyprus, Greece and Sweden, students had limited exposure to the application or received minimal guidance, which may have affected their ability to evaluate the its educational value. In most implementations, SB was used primarily at home rather than during class time, a factor that may have limited students' understanding of its intended use. Additionally, in some cases, student feedback collected through the questionnaire may have been influenced by low engagement with the survey process itself, as suggested by teacher observations in Greece during data collection.

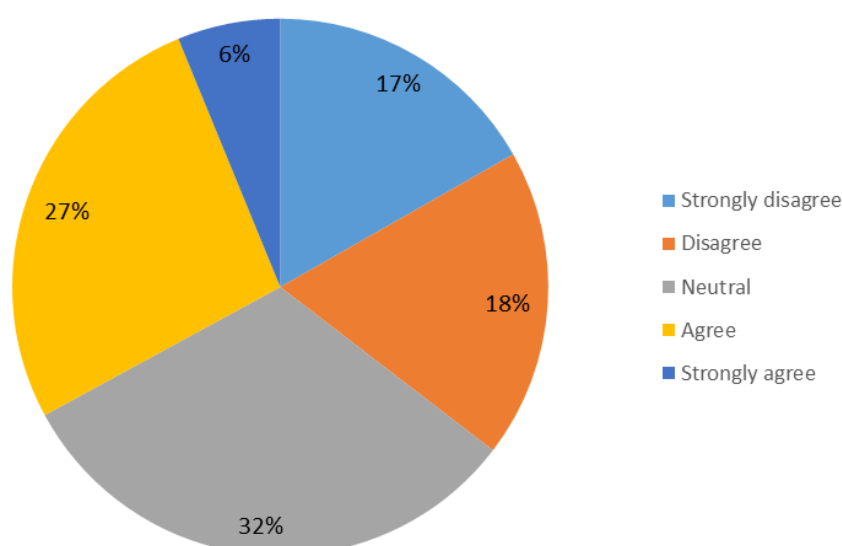


Figure 4.4: Student responses to Q4a – “Study Buddy helped me better understand the subject material”

- **Organization and Time Management**

Another item (Q4c) asked students whether SB helped them stay organized and manage their time effectively. Responses were again divided: 36 students (22%) agreed or strongly agreed, 65 (40%) remained neutral, and 64 (40%) disagreed or strongly disagreed (Figure 4.5). The relatively high percentage of neutral responses across items may indicate that students required more time or scaffolding to recognize and benefit fully from SB's potential academic support features. This question may not fully align with the actual use context, as, in most of the cases, SB was not integrated into a structured scenario involving systematic use of its organisational and planning features for specific learning tasks. Moreover, given that most implementations were short in duration and took place at the end of the school year, often constrained by other school commitments, students may not have had sufficient opportunity

to develop consistent usage habits or fully explore the organisational functions of the application.

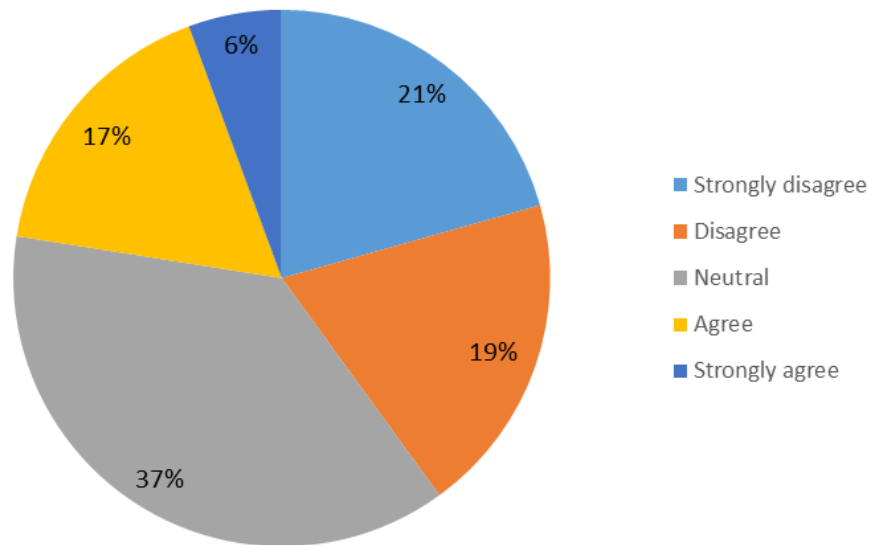


Figure 4.5: Student responses to Q4c – “Study Buddy helped me stay organized and manage my time effectively”

- **Confidence in Completing Assignments and Tests**

A third item (Q4d) examined students' confidence in completing academic assignments and tests after using SB. Only 36 students (22%) reported agreement or strong agreement, while 60 (37%) were neutral and 64 (40%) expressed disagreement or strong disagreement (Figure 4.6). While a portion of students reported increased confidence, overall perceptions of SB's impact on academic self-efficacy appeared limited, likely influenced by the fact that, in many cases, its use lacked sufficient structure for out-of-classroom use to support confidence-building. As with the previous question, sporadic use and a lack of alignment with specific assessment-related tasks may have constrained the application's perceived impact in this area.

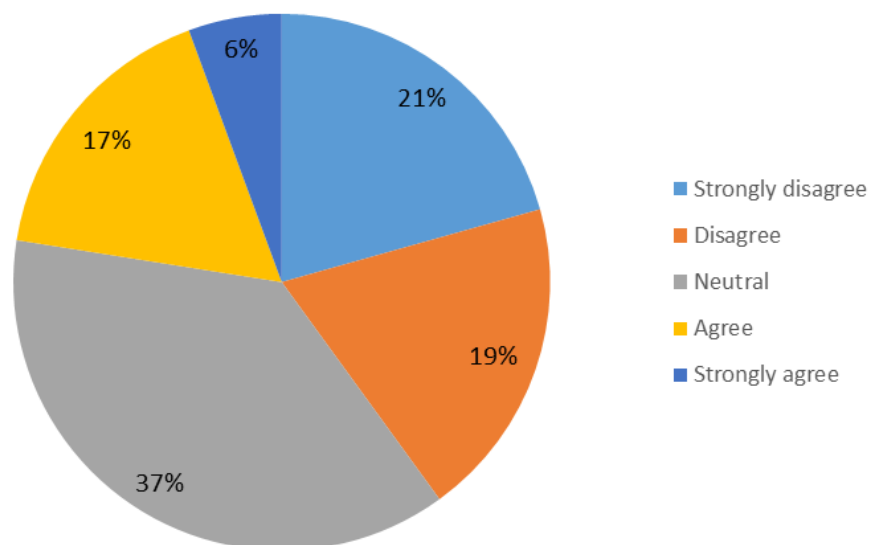


Figure 4.6: Student responses to Q4d – “Study Buddy improved my confidence in completing assignments and tests”

Overall, the quantitative data gathered from students’ responses to closed-ended survey questions in Cyprus, Greece, and Sweden revealed a mixed picture of SB’s perceived academic support, with many students expressing neutral or uncertain views, likely due to limited exposure and usage, minimal guidance, and lack of structured and prolonged integration. It should also be noted that, although a substantial number of students from Cyprus, Greece, and Sweden completed the survey (161 students), only a subset of these students (87 in total) were engaged with the application during the pilot implementations (section 4.3.1). This suggests that some responses may reflect limited exposure, based only on initial training or pre-implementation introductory sessions and demonstrations rather than sustained use. This context is essential for interpreting the large proportion of neutral responses and helps explain the relatively modest perceptions of academic impact. It also highlights the importance of triangulating data sources. In this light, the analysis of qualitative data in the following sections as well as the analysis of log files of actual application usage (Section 4.3) help contextualise and further interpret these survey findings and offer valuable complementary evidence to support a more multi-dimensional interpretation of SB’s educational contribution.

4.1.7.1.2. Student Feedback from Open-Ended Survey Responses (Cyprus, Greece, Sweden)

In open-ended responses, many students described SB as helpful in enhancing their academic performance, particularly in comprehension, preparation, and study support. Several students appreciated the clarity of explanations, the ease of access to summaries, and the application’s role in helping them prepare for tests. Specific references included that SB “helped with test preparation,” “summarized the key content,” and “explained things I didn’t understand in a simple way.” Some students also found the chatbot format engaging, noting that it made learning “more enjoyable” or helped them “explore information in different ways.”

Greek and Cypriot students emphasized that SB helped them “understand concepts more easily,” “answer questions immediately,” and “review history content,” with some describing it as “useful,” “smart,” or “perfect for studying.” Others appreciated the fact that they could “ask questions in different ways” and still receive answers that clarified what they had learned. A few noted that the application could support both individual study and group projects, especially in subjects requiring memorization.

Responses from Swedish students echoed similar themes. Many highlighted SB’s usefulness in summarising and explaining content, as well as supporting their ability to check their understanding. The chat-based format was frequently appreciated for providing tailored explanations of difficult terms or concepts, and for helping them gain clarity on areas they found confusing.

However, the qualitative data also highlighted several challenges. Some students reported that SB occasionally failed to understand their queries, delivered overly general or inaccurate answers, or lacked alignment with their textbooks. Technical issues were also noted, including difficulties with login, slow response times, and restricted content. A few students felt that SB did not help them retain information or that it failed to meet their learning needs.

4.1.7.1.3. Student Perspectives from Interviews in Ireland

In Ireland, additional insights were gathered through interviews with 24 students from the experimental group, and written reflections, providing a more personal perspective on SB's academic value. Following the implementation phase, students were invited to respond to the following questions:

- What did you like or dislike about using Study Buddy for your learning?
- What did you think of learning with Study Buddy compared to using a PowerPoint or worksheet?

These questions aimed to gather students' reflections on how SB supported their understanding, efficiency, organization, confidence, and enjoyment, and to explore whether the application offered a more engaging or effective learning experience compared to traditional instructional methods.

In response to these questions, many students emphasized that SB enhanced their understanding of complex topics, especially in History and Science. For example, one student shared, *"It helped me understand big words in history... the Explain tool made it easier for me."* Others appreciated how the application encouraged critical thinking: *"The AI was like having another teacher! It asked me questions and made me think about what I already knew."*

Several students highlighted specific features that supported academic progress. Tools like "Summarizer" and "Self-assessment quiz" were praised for facilitating research and test preparation. As one student noted, *"I used the summarise tool a lot... it helped me read less and still get what I needed for my project."* Another commented, *"I used the AI to make flashcards for my science test."* Others mentioned using the speech-to-text feature when tired or stressed, appreciating the ability to interact without needing to type: *"I could use the voice tool to ask questions when I felt too tired to type. That helped me a lot!"*

SB's ability to scaffold learning by giving hints rather than direct answers was positively received: *"When I didn't know an answer, the AI didn't just give it to me – it gave hints. This made me feel like I could figure things out on my own."* Students also appreciated the step-by-step guidance and the way SB explained why answers were right or wrong, helping them learn from their mistakes (*"It checked my answers and explained why I was right or wrong. This helped me learn from my mistakes"*), *"It felt like a teacher sitting next to me and showing me each step."*

However, some students expressed a need for greater accessibility and multimodal support. Learners with reading difficulties or attention challenges requested features such as more visuals, and gamified elements. For instance, one student explained, *"It was mostly words... I'd remember more if there were more ways to learn."* and another stated *"Sometimes the AI gave too much text. Adding more videos or audio would help kids like me who struggle with reading."*

Students also commented on how SB made learning more interactive and personalised, often contrasting it with traditional worksheets or PowerPoints. One student reflected, *"It's better because you can ask anything and not have to sit there for ages."* The ability to work at their own pace was particularly appreciated: *"you can work at your own pace and you don't have to rush cause other people are waiting and so is the teacher cause you're not as fast as others. Another benefit is you can ask more questions"* and *"I liked it because it was quick and easy."* Several described it as more enjoyable or motivating, stating: *"It even knew some funny jokes about math!"*

Some students also mentioned the tool “Chronological ordering of events” as particularly helpful for visualising content in History: *“AI4EDU had so many cool timelines that helped me see events in order. It made things less confusing.”*

Despite some usability issues and design suggestions, a recurring theme was that SB made learning more interactive and motivating. As one student summed up, *“It felt like a friend helping me,”* while another reflected, *“It explained things in simpler words when I got stuck – like a teacher sitting next to me.”* These voices highlight the importance of both pedagogical depth and user experience in maximizing the educational value of AI tools.

These reflections illuminate the diverse ways in which SB contributed to students’ understanding and self-regulated learning and reinforce its perceived educational value in real-world classroom contexts. Importantly, these qualitative insights are supported by usage data from the pilot implementations, as presented in Section 4.3, which provides additional evidence on how and to what extent students engaged with the application’s key features.

4.1.7.2. Teachers’ Reflections on Students’ Academic Performance

Teachers who participated in the pilot implementations were invited to share their impressions of the AI applications’ impact on students’ learning. Feedback on SB’s academic value was gathered through post-implementation questionnaires administered in Cyprus, Greece and Sweden, as well as through teacher interviews and observations collected in Ireland. These sources provided insights into students’ academic performance, and independent learning behaviors. However, many observations were shaped by contextual constraints, such as the short duration of the pilots, overlapping school commitments, and limited opportunities for meaningful or sustained use of the application.

Across the participating countries, several teachers reported observable benefits when SB was used to support academic tasks. In Cyprus, one teacher noted an improvement in students’ project work following the integration of SB, while in Sweden, students were seen using the application to explore questions related to historical periods and literature. Teachers acknowledged SB’s potential to foster independent learning; one Swedish teacher remarked that although students still asked the teacher for help, they were increasingly encouraged to consult SB, gradually developing more autonomous study habits.

Despite these positive accounts, most teachers reported that the restricted implementation period made it difficult to draw clear conclusions about the application’s academic impact. In both Cyprus and Sweden, the application was not used frequently or systematically enough to enable meaningful comparisons between experimental and control groups or to detect significant changes in learning outcomes.

In addition to time constraints, other challenges were noted. These included difficulties incorporating SB into regular classroom routines, students’ limited familiarity with how to use the application effectively, and technical issues such as login problems or poor internet connectivity. Nonetheless, no incidents of student misuse were reported, and teachers generally expressed interest in the tool’s future use, especially if its interface and classroom integration were improved.

Rich insights were obtained through teacher interviews and classroom observations in Ireland gathered by three teachers who participated in the first phase of pilots. Teachers of the experimental group reported several encouraging changes in student behavior and academic engagement following the introduction of SB. They observed that the application encouraged more proactive learning behaviors and deeper questioning. One teacher noted that students became more independent, initiating discussions and proposing new ideas,

particularly during sustainability-themed lessons. SB's instant feedback was seen as a valuable feature, helping students reflect on their work and identify areas for improvement without waiting for teacher input.

Classroom observations supported these findings. Compared to their peers in the control group, students in the experimental classes demonstrated enhanced comprehension and participation during interactive tasks. Analysis from the SCOT and NOISE observation tools highlighted stronger critical thinking, greater student agency, and improved alignment with learning goals in the SB-supported classrooms. Teachers also emphasized SB's value in supporting differentiated instruction, as its personalized feedback allowed students to progress at their own pace.

However, Irish teachers also acknowledged some limitations. Some students initially struggled to navigate the application's interface, and there were signs of overreliance on SB in tasks requiring independent reasoning. Teachers emphasized the need for clearer onboarding and better integration into the curriculum to fully unlock SB's potential, particularly for promoting inquiry-based and reflective learning.

In sum, teachers' reflections across the pilot sites pointed to promising uses of SB for supporting students' academic development. While clear impacts on performance were difficult to quantify in such a short implementation window, the application was generally seen as having educational value, particularly in fostering independent learning, critical thinking, and student engagement, provided that sufficient training, time, and structural support are in place for effective integration.

4.2. Motivation and Engagement

This section examines the impact of the AI4EDU applications on students' motivation and engagement, drawing on data from the Motivation and Engagement Survey (MES), post-implementation student questionnaires, teacher feedback, and classroom observations. The analysis aims to determine whether and how the use of SB contributed to students' engagement, as well as their overall motivation toward learning. Responses are analyzed across multiple pilot sites, highlighting patterns and variations in student experiences. The findings are presented through descriptive statistics, comparative tables, and selected examples that shed light on the ways AI-supported learning environments may shape students' attitudes, effort, and participation in the learning process.

4.2.1. Survey-Based Evaluation of Motivation and Engagement

To assess changes in students' motivation and engagement, MES was administered in three pilot implementations conducted in Cyprus and Greece. In each case, MES was distributed to both experimental and control groups at two points in time: before the introduction of the AI4EDU applications (pre-implementation) and immediately after the completion of pilot activities (post-implementation). This pre-post design enabled the identification of potential shifts in motivational and engagement profiles associated with the use of SB.

MES was adapted from the validated framework developed by Lee and Reeve (2012). It consists of 38 items, each rated on a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree), grouped into the following dimensions:

- Psychological Need Satisfaction (Items 1–10)
- Self-Efficacy (Items 11–15)
- Mastery Goals (Items 16–18)
- Behavioral Engagement (Items 19–23)

- Emotional Engagement (Items 24–28)
- Cognitive Engagement (Items 29–33)
- Agentic Engagement (Items 34–38)

For clarity and reporting purposes, the items were grouped into two broad categories:

- **Motivation (Items 1–18)**, capturing students’ internal drivers of learning, such as autonomy, self-efficacy, and mastery orientation.
- **Engagement (Items 19–38)**, capturing observable indicators of student participation and involvement, including behavioral, emotional, and cognitive engagement.

To ensure consistency in scoring direction, responses to negatively worded items (Items 4, 23, 28, 32, and 33) were reverse-coded so that higher scores uniformly indicated more favorable motivational and engagement profiles.

The analysis followed a three-step process:

- **Score Aggregation:** For each student, average scores for motivation and engagement were calculated at both time points (pre- and post-implementation). Indicative examples of individual-level results are shown in Table 4.19.
- **Comparative Analysis:** Aggregated scores were then compared across groups and conditions. Specifically, analyses explored:
 - **Within-group comparisons:** Pre- and post-implementation scores were compared within the experimental groups to assess changes in students’ motivation and engagement following the use of SB.
 - **Between-group comparisons:** Experimental and control groups were compared to evaluate the differential impact of the AI-supported learning environment relative to traditional instruction.
 - **Between-subject comparisons:** Differences were explored across subject to examine whether the application’s impact varied by disciplinary context.
 - **Demographic comparisons:** Patterns were analyzed based on gender to determine whether individual differences influenced the outcomes.

Table 4.19: Individual-Level Pre and Post Survey Scores

No	Subject	Group	Motivation pre	Motivation post	Engagement pre	Engagement post
1	Physics	Control	3,33	3,44	3,15	3,50
2	Physics	Experimental	3,06	3,25	3,78	3,60

4.2.1.1. Within Group Analysis

To evaluate the overall impact of SB on students’ motivation and engagement, a within-group analysis was conducted by aggregating data from all experimental groups across the three implementation cases (Physics, Chemistry, and History). The standardized format and consistent administration of MES enabled the integration of data across subject areas, allowing for a comprehensive assessment of changes over time.

For each student in the experimental groups, mean scores were calculated separately for the constructs of motivation (items 1–18) and engagement (items 19–38) at two time points: before (pre) and after (post) the implementation of the AI4EDU application. Descriptive statistics, including pre- and post-implementation means, are presented in Table 4.20, while Figure 4.7 provides a visual summary of the observed trends.

Table 4.20: Mean Pre- and Post-Implementation Scores for Motivation and Engagement (Experimental Group)

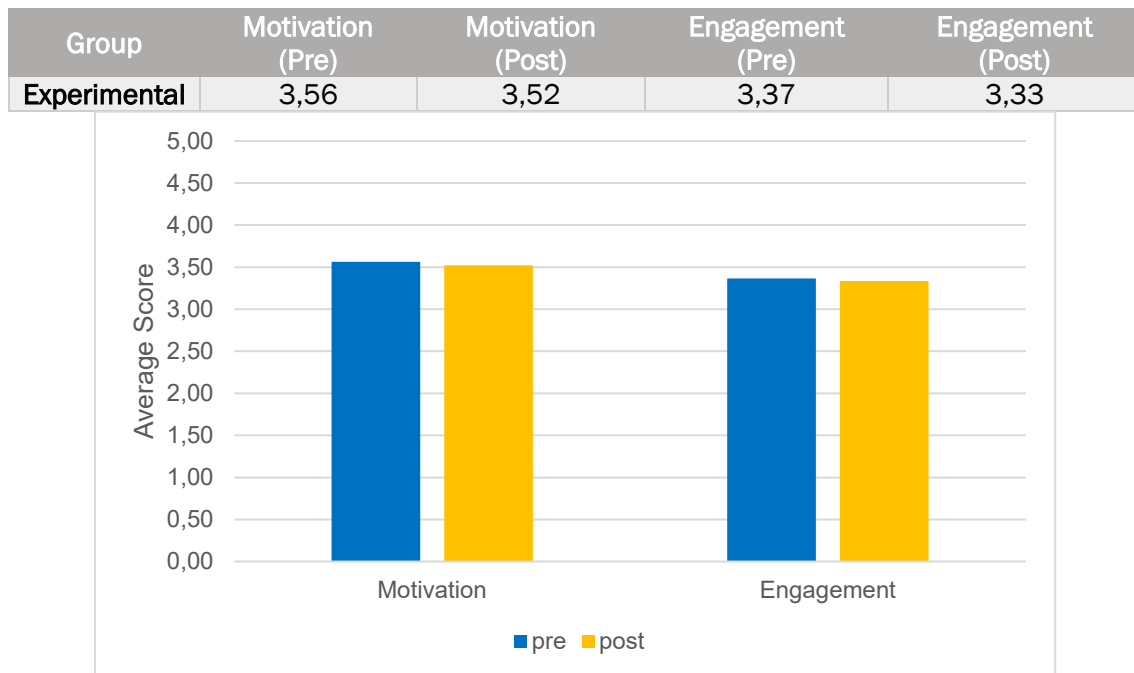


Figure 4.7: Pre- and Post-Implementation Mean Motivation and Engagement Scores

The results suggest that students in the experimental group maintained relatively stable levels of motivation and engagement throughout the implementation period. On average, only slight decreases were observed, indicating no substantial decline in students' attitudes or participation levels while using SB.

To determine whether these observed changes were statistically significant, a Shapiro–Wilk test was first conducted to assess the normality of score distributions for both constructs. As presented in Table 4.21, none of the distributions deviated significantly from normality, supporting the use of parametric testing. Subsequently, paired-samples t-tests were performed to compare students' pre- and post-implementation scores. As shown in Table 4.22, no statistically significant differences were found for either motivation or engagement.

Table 4.21: Shapiro-Wilk Test of Normality (Within Group Analysis)

Construct	Statistic	Degrees of freedom (df)	Significance (p)
Motivation (Pre)	0,989	52	0,917
Motivation (Post)	0,973	52	0,278
Engagement (Pre)	0,992	52	0,981
Engagement (Post)	0,986	52	0,804

Table 4.22: Paired-Samples t-Test Results (Within Group Analysis)

Test	t value	Significance (p)
Motivation (Pre) – Motivation (Post)	0,776	0,441
Engagement (Pre) – Engagement (Post)	0,639	0,526

These results indicate that the use of SB did not lead to statistically significant changes in students' motivation or engagement during the pilot period. However, the stable patterns

suggest that the application may have helped sustain students' learning dispositions, especially in contexts where engagement often declines over time.

4.2.1.2. Between Group Analysis

In addition to within-group comparisons, a between-group analysis was performed to examine whether the use of SB had a differential impact on students' motivation and engagement compared to traditional instruction. This analysis compared the aggregated mean scores of students in the experimental groups with those in the control groups across the three pilot cases. The objective was to determine whether students who engaged with the AI-supported learning environment experienced more favorable changes in motivation and engagement than their peers in the control group.

Table 4.23 presents the aggregated pre- and post-implementation mean scores for both constructs, while Figure 4.8 provides a visual representation of the trends observed across the two groups.

Table 4.23: Aggregated Mean Scores for Motivation and Engagement by Group

Group	N	Motivation (Pre)	Motivation (Post)	Engagement (Pre)	Engagement (Post)
Control	52	3,54	3,44	3,40	3,33
Experimental	52	3,56	3,52	3,37	3,33

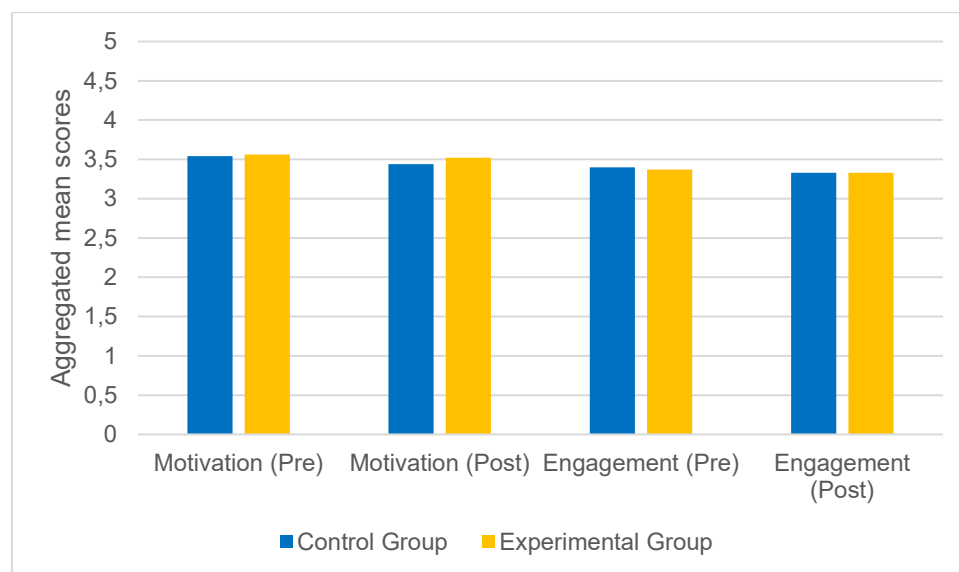


Figure 4.8: Comparison of Pre- and Post-Implementation Mean Scores for Motivation and Engagement by Group

The descriptive results suggest that both groups maintained relatively stable levels of motivation and engagement during the implementation period. While the control group showed a slight decline in both constructs ($-0,10$ in motivation, $-0,07$ in engagement), the experimental group exhibited only minimal decreases ($-0,04$ and $-0,03$, respectively). Although the differences between groups appear modest, they may indicate a small advantage in favor of the experimental group in sustaining student motivation and engagement during the intervention.

To determine whether these differences were statistically significant, a Shapiro–Wilk test was first conducted to assess the normality of score distributions for both constructs across the two groups. As shown in Table 4.24, none of the distributions deviated significantly from normality ($p > 0,05$), supporting the use of parametric testing.

Subsequently, independent samples t-tests were conducted to compare the two groups at both time points. As presented in Table 4.25, no statistically significant differences were observed between the groups in either construct at pre- or post-implementation.

Table 4.24: Shapiro-Wilk Test of Normality (Between Group Analysis)

Construct	Group	Statistic	Degrees of freedom (df)	Significance (p)
Motivation (Pre)	Control	0,992	52	0,979
	Experimental	0,989	52	0,917
Engagement (Pre)	Control	0,988	52	0,883
	Experimental	0,973	52	0,278
Motivation (Post)	Control	0,984	52	0,727
	Experimental	0,992	52	0,981
Engagement (Post)	Control	0,974	52	0,322
	Experimental	0,986	52	0,804

Table 4.25: Independent Samples t-Test Results (Between Group Analysis)

Construct	t value	Significance (p)
Motivation (Pre)	-0,245	0,807
Engagement (Pre)	0,323	0,747
Motivation (Post)	-0,768	0,444
Engagement (Post)	-0,030	0,488

These findings suggest that the use of SB did not result in statistically significant differences in students' motivation or engagement compared to traditional instruction. Nonetheless, the slightly smaller declines observed in the experimental group may be interpreted as a positive indication that AI-supported environments can help maintain student learning dispositions under certain conditions. Further research with larger samples and longer intervention periods could help clarify these trends.

4.2.1.3. Between-Subject Analysis

To explore whether the impact of SB varied depending on the subject area, a between-subject analysis was conducted. This comparison focused on the experimental groups only, examining differences in students' motivation and engagement scores across the three pilot implementation cases: Physics, Chemistry, and History.

Mean pre- and post-implementation scores were calculated separately for each subject area and are presented in Table 4.26. Figure 4.9 provides a comparative visual representation of the mean scores across the three subjects.

Table 4.26: Mean Pre- and Post-Implementation Scores for Experimental Groups by Subject

Subject	N	Motivation (Pre)	Motivation (Post)	Engagement (Pre)	Engagement (Post)
Physics	21	3,48	3,47	3,15	3,31
Chemistry	17	3,64	3,68	3,66	3,46
History	14	3,58	3,42	3,35	3,22

These findings reveal modest differences in how students' motivation and engagement evolved depending on the subject. Motivation scores slightly increased in Chemistry, remained nearly stable in Physics, and declined in History. Engagement scores, on the other hand, improved in Physics, but decreased in Chemistry and History.

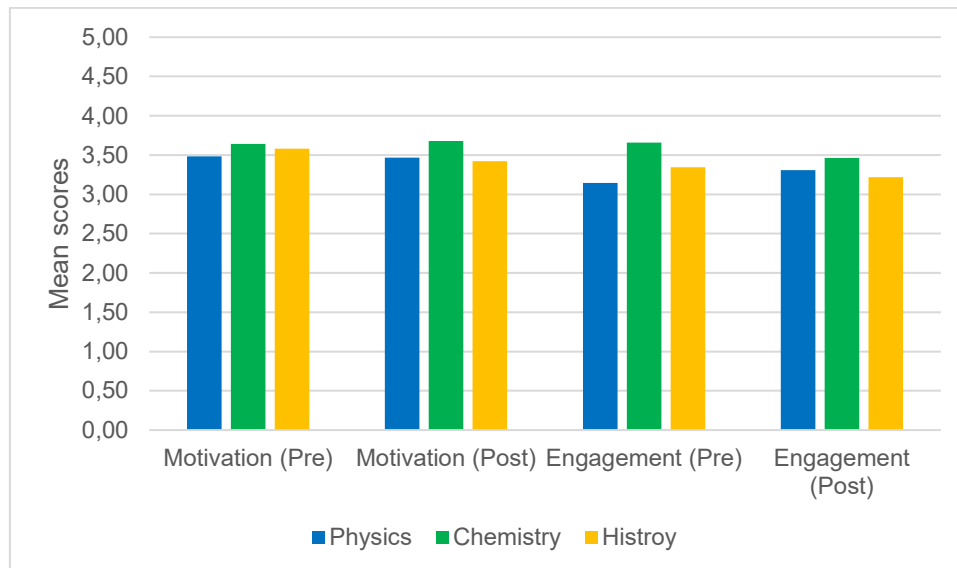


Figure 4.9: Mean Scores in Motivation and Engagement Across Subject Domains

To determine whether these differences were statistically significant, a Shapiro–Wilk test was first conducted to assess the normality of score distributions for each construct across the three subjects. As shown in Table 4.27, all distributions met the normality assumption ($p > 0,05$), supporting the use of parametric testing.

Table 4.27: Shapiro-Wilk Test of Normality (Between-Subject Analysis)

Construct	Group	Statistic	Degrees of freedom (df)	Significance (p)
Motivation (Pre)	Physics	0,987	21	0,988
	Chemistry	0,973	17	0,867
	History	0,951	14	0,572
Engagement (Pre)	Physics	0,944	21	0,266
	Chemistry	0,947	17	0,412
	History	0,894	14	0,094
Motivation (Post)	Physics	0,943	21	0,252
	Chemistry	0,983	17	0,978
	History	0,973	14	0,916
Engagement (Post)	Physics	0,990	21	0,998
	Chemistry	0,965	17	0,720
	History	0,980	14	0,977

Subsequently, one-way ANOVA was conducted to compare pre- and post-implementation scores across the three subjects. As shown in Table 4.28, no statistically significant differences were observed for motivation scores. However, a significant difference was observed in pre-implementation engagement scores, suggesting that initial engagement levels varied by subject. Post-implementation engagement scores did not differ significantly.

Table 4.28: One-Way ANOVA Results (Between-Subject Analysis)

Construct	F	Significance (p)
-----------	---	------------------

Motivation (Pre)	0,452	0,639
Engagement (Pre)	4,835	0,012
Motivation (Post)	1,054	0,356
Engagement (Post)	0,795	0,457

In summary, although descriptive trends point to small variations across subjects, particularly with Chemistry showing a slight motivational increase and Physics a gain in engagement, these differences were not statistically significant at post-test. This suggests that the educational impact of SB on students' motivation and engagement was broadly consistent across the three disciplinary contexts.

4.2.1.4. Demographic Analysis

To explore whether the impact of SB varied across demographic subgroups, a gender-based analysis was conducted focusing on students in the experimental groups. The objective was to assess whether the AI-supported learning environment influenced boys and girls differently in terms of motivation and engagement.

Mean pre- and post-implementation scores were calculated separately for female and male students across both constructs. As shown in Table 4.29, female students exhibited a slight increase in motivation (+0,04) and maintained stable engagement levels, while male students experienced a small decrease in both motivation (-0,12) and engagement (-0,07).

Table 4.29: Mean Motivation and Engagement Scores by Gender (Experimental Group)

Gender	N	Motivation (Pre)	Motivation (Post)	Engagement (Pre)	Engagement (Post)
Female	26	3,51	3,55	3,37	3,37
Male	26	3,62	3,50	3,37	3,30

To determine whether these differences were statistically significant, Shapiro-Wilk tests were first performed to assess the normality of distributions for each construct by gender. As shown in Table 4.30, all distributions met the assumption of normality ($p > 0,05$), justifying the use of parametric testing.

Table 4.30: Shapiro-Wilk Test of Normality (Demographic Analysis)

Construct	Group	Statistic	Degrees of freedom (df)	Significance (p)
Motivation (Pre)	Female	0,973	26	0,711
	Male	0,970	26	0,618
Engagement (Pre)	Female	0,958	26	0,347
	Male	0,929	26	0,072
Motivation (Post)	Female	0,977	26	0,797
	Male	0,975	26	0,761
Engagement (Post)	Female	0,983	26	0,930
	Male	0,976	26	0,769

Subsequently, independent samples t-tests were conducted to examine differences between girls and boys. As shown in Table 4.31, no statistically significant differences were found between genders at either time point for either construct ($p > 0,05$).

Table 4.31: Independent Samples t-Test Results (Demographic Analysis)

Construct	t value	Significance (p)
Motivation (Pre)	-0,771	0,444
Engagement (Pre)	0,000	1,000

Motivation (Post)	0,314	0,755
Engagement (Post)	0,471	0,640

These findings suggest that SB was similarly effective for both female and male students in terms of maintaining motivation and engagement. Although small numerical differences were observed, they were not statistically meaningful. The results indicate that the application did not favour one gender over the other in its motivational or engagement impact.

4.2.2. Student and Teacher Reflections on Motivation and Engagement

To complement the quantitative findings on motivation and engagement, feedback was collected from students and teachers to gain deeper insights into how the SB application influenced students' motivation and engagement. This section draws on responses from post-implementation questionnaires and interviews with both students and teachers, as well as classroom observation notes. These complementary sources provide a nuanced understanding of students' learning behaviors, attitudes, and emotional responses throughout the intervention period. The findings are presented in two subsections: the first focuses on students' self-reported experiences with the application, while the second highlights teachers' observations and perspectives regarding changes in student motivation and engagement.

4.2.2.1. Student Feedback on Motivation and Engagement

In addition to the MES, student perspectives on motivation and engagement were further explored through the post-implementation questionnaires administered after using SB during the pilot phase. A key item in the questionnaire asked students to indicate their level of agreement with the statement: "Study Buddy made studying more engaging and enjoyable." Responses were recorded using a 5-point Likert scale ranging from Strongly Disagree (1) to Strongly Agree (5).

Out of 161 total responses collected from Cyprus, Greece, and Sweden, 19 students (12%) strongly disagreed, 54 (34%) disagreed, 104 (65%) selected neutral or positive responses, with 36 students (22%) agreeing and 19 students (12%) strongly agreeing. These results suggest a moderately positive reception: while a significant proportion of students remained neutral or unconvinced, more than one-third perceived SB as an application that contributed to a more enjoyable and engaging learning experience.

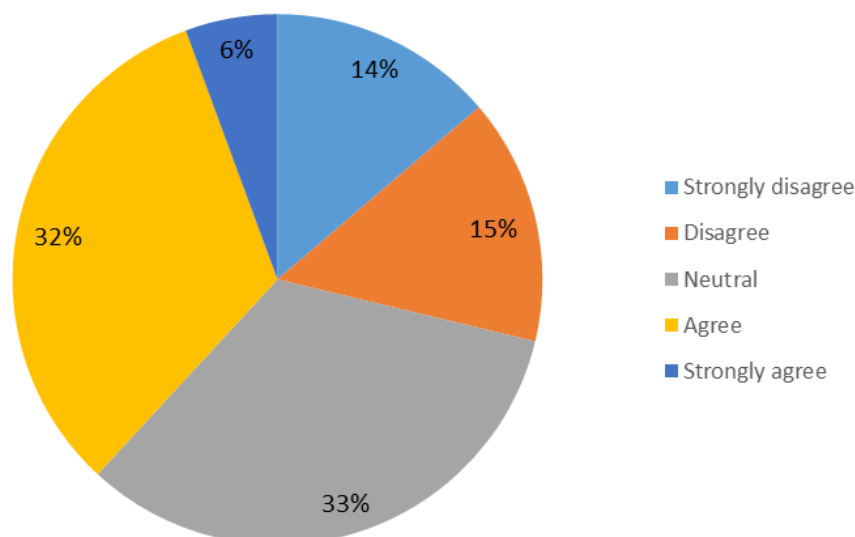


Figure 4.10: Student responses to Q4b – “Study Buddy made studying more engaging and enjoyable”

These results suggest that, while a substantial portion of students perceived SB as a motivating and engaging learning tool, a significant share remained indifferent or unpersuaded about its benefits in this area. These student-reported perceptions align with the findings from the MES analysis, which showed generally stable but non-significant changes in motivation and engagement following the intervention.

To complement these quantitative results, qualitative data from student interviews, open-ended questionnaire responses, and structured classroom observation tools (SCOT and NOISE) provided richer insights into students’ motivational and emotional experiences.

Many students in the Irish pilot reported that SB helped them feel more engaged and autonomous in their learning. They appreciated the immediacy of feedback and the opportunity to explore content at their own pace. As one student noted, *“I think it is better because you can work at your own pace, and you don’t have to rush because other people are waiting.”*

SB also appeared to foster a more enjoyable and stimulating learning environment. Students described SB as “very interesting to learn” and praised its interactivity. One student reflected, *“It gives you way more info and makes you think more about how to help the earth,”* suggesting deeper engagement with the content, particularly in sustainability-related lessons. Others reported feeling “proud” when they received hints instead of answers helping them to find the answer themselves, highlighting a positive shift in learning attitudes.

Emotional engagement was also evident in remarks such as *“It made me feel like I had help, even if the teacher was busy,”* and *“When I typed something wrong, it didn’t say ‘wrong!’, it gave me help. That made me feel brave.”* Students highlighted how the non-judgmental nature of the feedback encouraged risk-taking and boosted confidence.

At the same time, some students pointed to challenges that may have tempered their overall motivational response. A common concern involved the initial difficulty of navigating the application: *“It was a bit confusing at first, but then it got better.”* Others mentioned that too much reliance on the AI could lead to less critical thinking: *“It’s better than writing on a sheet because AI tells us the answers.”*

In terms of opportunities for improvement, students suggested the inclusion of more dynamic and gamified elements: *“It would be fun if there were more challenges.”* There was also a call for enhanced accessibility, such as visual aids.

Overall, these findings suggest that while quantitative measures reflected moderate impact, qualitative data reveals that many students experienced SB as a motivating and empowering application. Its ability to support self-paced learning, increase confidence, and promote a sense of agency was especially valued, provided that sufficient orientation and support were in place during the implementation. These positive effects were particularly evident in the Irish implementation, which was characterized by full in-class integration, supported in situ by the engagement of AI4EDU team members, and highly engaged teachers with prior exposure to the application. The younger age group of Irish students may also have contributed to a more enthusiastic reception, as students responded well to interactive and supportive feedback mechanisms within a structured learning environment, in contrast to older secondary school students with more demanding schedules and duties.

4.2.2.2. Teacher Reflections on Students' Motivation and Engagement

Teachers participating in the pilot implementations were asked to share their impressions of how the SB application influenced student motivation and engagement. Post-implementation questionnaires were administered in Cyprus, Greece, and Sweden, while additional insights were obtained from teacher interviews and classroom observations conducted in Ireland. These data sources provided a well-rounded view of how SB shaped students' interest, participation, and emotional involvement during learning.

In Cyprus, some teachers noted an initial excitement among students when SB was first introduced. However, this enthusiasm tended to diminish once SB became associated with formal assignments. As one teacher explained, *"At first they were excited, but their attitude changed when assignments were given"*. Other teachers reported no significant change in student motivation or classroom engagement. Similarly, Swedish teachers stated that the short implementation period limited their ability to draw conclusions: *"Since the project coincided with absences and other activities, we didn't use Study Buddy long enough to draw reliable conclusions"*.

Several challenges were reported that may have limited SB's motivational impact. These included difficulty integrating the application into classroom routines, lack of familiarity with its features, and technical issues such as login problems and unstable internet access. Nonetheless, there were no incidents of misuse, and teachers expressed interest in further exploring the application's potential once these barriers were addressed.

Teacher interviews and classroom observations in Ireland offered additional insights. Teachers from the experimental group described SB as a useful tool for fostering deeper student engagement. They observed that the application promoted critical thinking and encouraged students to take initiative during lessons. One teacher commented that *"The students using Study Buddy were much more proactive. They asked deeper questions and even suggested new sustainability projects for the school."* Another highlighted the application's immediate feedback function, noting that it helped students identify areas for improvement independently, without waiting for the teacher's input.

Observation data from SCOT and NOISE frameworks reinforced these findings. AI-supported classrooms showed greater signs of student agency, reflective engagement, and alignment with lesson objectives, particularly during sustainability-themed activities. Differentiated instruction was also more effectively implemented, as SB enabled students to work at their own pace and receive targeted support based on their needs.

However, Irish teachers also pointed out some limitations. A few students initially struggled to navigate the application, and there were instances where students became overly reliant on SB, especially in situations requiring personal reflection or independent reasoning. Teachers emphasized that a more structured onboarding process and curriculum integration would be essential to maximize SB's benefits for motivation and engagement.

In sum, while the overall motivational impact of SB varied across the pilot sites, partly due to contextual and technical constraints, teachers generally recognized its potential to enhance student engagement, foster inquiry, and support differentiated learning when implemented more systematically.

4.2.2.2.1. Case study: Inclusive AI integration for students with Special Educational Needs (Ireland)

- Overview and context

As outlined in section 3.3, St. Thomas' Senior National School in Jobstown, Ireland, hosted a concentrated pilot involving approximately 72 students from 5th and 6th class, including both control and AI-enhanced experimental groups. A committee of 8 educators was involved, including three classroom teachers, an EAL specialist, SEN teacher, a senior member of the school management team, and two representatives from the DEC, members of the AI4EDU team, who were present at the whole pilot's implementation period. As a DEIS school, St. Thomas Senior National School supports students from socioeconomically disadvantaged backgrounds, including students with SEN, students with EAL, such as Ukrainian immigrant pupils, as well as children from Roma/Traveller backgrounds. The rich diversity in learning profiles makes the school a suitable setting for exploring inclusive AI interventions.

The implementation was framed by an organisational psychology perspective, positioning the classroom as a socio-cognitive system shaped by the interplay of tools, tasks, and social dynamics. This lens emphasized motivational climates, teacher-student interaction patterns, and peer learning ecologies. The pilot sought to observe how AI technologies might influence not just individual engagement, but also broader relational and systemic learning dynamics, especially in inclusive and under-resourced contexts. This emphasis to organisational perspectives is aligned with the objectives of Work Package 7, led by DEC, which aims to produce evidence-based policy recommendations and practical implementation guidelines for AI integration in education across diverse European contexts.

Methodologically, the intervention followed a flexible, mixed qualitative design. In-class observations were guided by SCOT and NOISE tools to capture behavioural, affective, and agentic dimensions of engagement. Teacher reflections were collected through structured interviews conducted post-intervention, and student perspectives were gathered via informal oral feedback. In addition, an organisational psychology questionnaire was administered to participating teachers. This instrument focused on perceived changes in classroom atmosphere, workload, and instructional autonomy, supporting a multi-level analysis of AI's systemic impact on both teaching and learning dynamics. While not designed for experimental generalization, the pilot prioritized contextual sensitivity, rich description, and multiple perspectives to surface meaningful patterns related to engagement, inclusion, and emerging AI-supported practices.

- **Key findings**

The findings from the pilot at St. Thomas' Senior National School reveal a multi-layered and encouraging impact of AI integration on learner engagement, inclusion, and classroom dynamics, particularly within the context of a DEIS primary school.

Regarding, student engagement and learning gains, in the experimental group, observers documented a 38% increase in engagement². Students demonstrated heightened cognitive and emotional involvement with the sustainability curriculum. SB's real-time feedback mechanisms were particularly effective in fostering metacognitive reflection and language development, as evidenced by richer use of academic vocabulary and greater confidence in oral and written communication. Teachers noted that even previously disengaged students were more active, with several stating that SB helped them "find their voice" and express complex ideas using terms such as "ecosystem" and "citizenship".

² Quantified results of students' progress are derived from teacher surveys and structured classroom observations and were further corroborated by internal school assessment records, within the constraints of personal data protection and school internal protocols.

Students with SEN, those with EAL, and Roma or Traveller backgrounds responded especially well to the personalisation and encouragement offered by AI applications. Enhanced attendance and reduced behavioural incidents were reported during the intervention period. Notably, Ukrainian immigrant students showed a 41% improvement in English language acquisition, and 67% of teachers reported higher engagement levels among marginalised learners.

Additionally, the organisational psychology component revealed a strong link between school culture and the success of AI integration. Teachers in the experimental group reported high levels of psychological safety (4.8/5), indicating a climate conducive to experimentation, shared leadership, and reflective practice. Observational and survey data from organisational psychology questionnaire demonstrated that when teachers felt safe to take risks and co-design implementation processes, the adoption of SB became more effective and rewarding.

Motivational shifts in both teachers and students were demonstrated. Autonomy was supported through SB's and TM's customisation options; competence was reinforced via adaptive feedback; and relatedness was strengthened as students collaborated around shared AI-supported tasks. Teachers reported that the tools allowed them to shift from a direct instructional role to one of facilitation and mentoring.

The organisational survey further illuminated key enablers and inhibitors of implementation. Schools like St. Thomas', with pre-existing collaborative norms and distributed leadership structures, exhibited greater agility in integrating AI. Teachers described real-time iteration loops, dynamic lesson adjustment, and a shared sense of ownership. The pilot acted as a catalyst for emerging communities of practice and reflective professional cultures. Crucially, these shifts were not confined to individual practice. Teachers reported enhanced institutional capacity for innovation, supported by a psychologically safe environment and aligned leadership. As one teacher expressed, *"We grew as a team because the tools encouraged dialogue, not just with students but with each other."*

4.2.3. Concluding remarks

In conclusion, this analysis provides a comprehensive and balanced account of SB's impact on student motivation and engagement, grounded in both statistical findings and rich qualitative evidence. Quantitative analyses across Cyprus, Greece, and Sweden indicated no statistically significant changes, but demonstrated the application's ability to sustain learners' motivational and engagement levels. These findings are reinforced by qualitative data from open-ended responses, teacher interviews, and structured classroom observation protocols, which revealed that students, especially in the Irish pilot, experienced heightened emotional and cognitive engagement, increased autonomy, and a more personalized, responsive learning environment. Overall, the study demonstrates that although short-term implementation may limit measurable impact, under mature conditions of integration, prior exposure, and teacher readiness, SB shows strong potential to enhance students' motivation in ways that reflect current educational priorities, such as fostering learner autonomy, promoting engagement through personalized feedback, and encouraging deeper cognitive involvement in learning tasks. Moreover, the importance of organisational enablers such as psychological safety, collaborative leadership, and inclusive classroom practices is underscored. These insights highlight the need to view AI integration not only through a technological lens but also through systemic and cultural frameworks that prioritise equity, agency, and teacher empowerment.

In general, both quantitative and qualitative analysis offer empirical grounding for the development of actionable implementation guidelines and policy recommendations,

emphasizing the importance of adequate onboarding, structural embedding of AI tools, and teacher facilitation as critical enablers of meaningful motivational impact.

4.3. Application Usage

This section presents an overview of how the two AI4EDU applications were used during the second pilot cycle. Drawing on log file data, the analysis explores patterns of use, levels of engagement, and user experiences with the applications across participating countries. Usage metrics, including the number of unique users, chat frequency and duration, and feature popularity are examined to shed light on the scale and nature of application use.

4.3.1. Study Buddy Log File Analysis

This subsection focuses on student interactions with SB. Drawing on log file data collected during the second pilot cycle, the analysis presents a quantitative overview of how students engaged with the application across all countries. Specifically, it examines the number of unique student users, the frequency and duration of chat sessions, and general patterns of engagement. These data offer valuable insights into how students integrated SB into their study routines and the extent to which they explored its capabilities in educational settings.

A total of 130 students actively engaged with SB during the pilot phase. Figure 4.11 displays the **distribution of unique student users** across the four participating countries. Cyprus, Greece, and Ireland recorded approximately the same number of users, while Sweden had limited student participation.

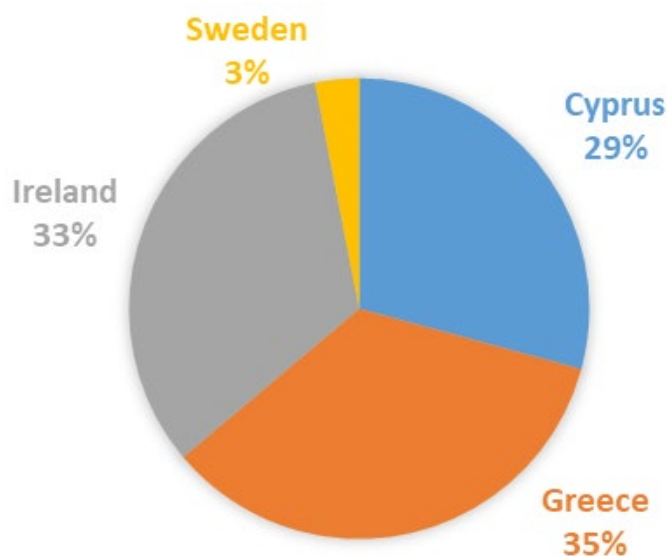


Figure 4.11: Distribution of Unique Student Users per Country

To gain a better understanding of user engagement, the number of times each student interacted with SB was also analyzed. As shown in Figure 4.12, a substantial proportion of students (97 out of 130, or 75%) used the application between one and five times, indicating exploratory or occasional use. Another 16 students (12%) engaged with the tool between 5 and 10 times, while a more engaged subgroup of 17 students (13%) initiated more than 10 chats. Usage in Cyprus, Greece, and Sweden was primarily concentrated in the lower interaction range. In contrast, students in Ireland demonstrated deeper engagement, with several users initiating a notably high number of sessions. This kind of intense usage points to potential use cases for highly autonomous or motivated learners, as was shown in the

analysis of Irish students' responses in interviews (4.1.7.1.3.), their teachers' reflections on their academic performance (4.1.7.2) and qualitative data about Irish students' motivation and engagement (4.2.2).

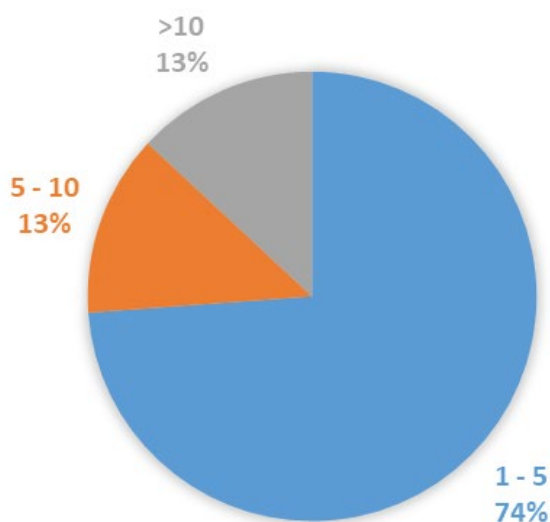


Figure 4.12 Distribution of Number of Study Buddy Interactions per User

Chat sessions were further analyzed based on their duration and classified into short (0–10 minutes), medium (10–60 minutes), and long (>60 minutes) categories. As illustrated in Figure 4.13, short sessions were the most common, accounting for 467 out of 672 total chat sessions (70%). Medium-length sessions represented 29% (192 chats), while 2% (13 chats) lasted over an hour. Ireland stood out with the highest number of sessions overall (420), including the largest number of medium-duration chats (143), suggesting more sustained and engaged use of the application. Greece and Cyprus showed moderate activity, while usage in Sweden was limited. The predominance of short-duration sessions indicates that students most commonly used SB for brief, task-specific purposes such as clarifying concepts or reviewing material. However, the presence of longer sessions highlights that a subset of students utilized SB for more extensive and focused study periods.

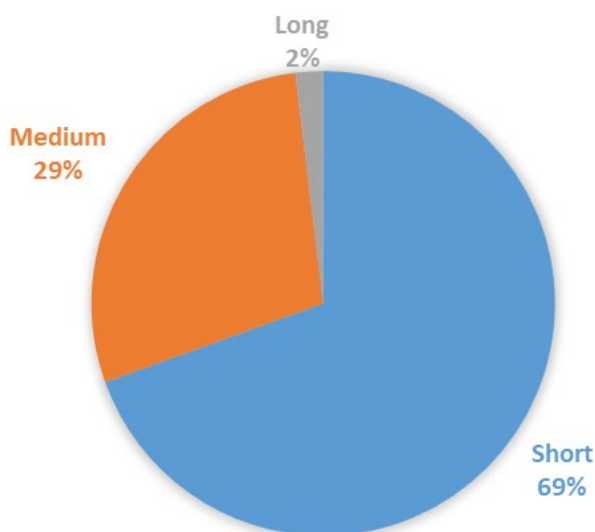


Figure 4.13: Distribution of Study Buddy Interactions by Duration

In addition to overall engagement patterns, the log file data allowed for an examination of which SB tools students used most frequently during the pilot implementation. These tools, embedded within the application, were designed to support various learning objectives, ranging from content comprehension to assessment preparation, and offer insight into how students chose to interact with AI-driven educational support.

Figure 4.14 presents the number of times each tool within the SB application was used by students. As shown, the most frequently used feature by far was the general “Chat with Study Buddy” function, with a total of 273 recorded uses. In each participating country, this tool emerged as the most popular, suggesting a consistent preference among students. This feature enables open-ended, conversational interaction with the AI assistant, supporting quick questions, conceptual clarifications, and general academic guidance. Its widespread use likely reflects students’ inclination toward intuitive, flexible tools and may be influenced by its resemblance to familiar AI technologies such as ChatGPT.

The second most popular tool was “Explain Term” (73 uses), demonstrating that many students leveraged SB to clarify subject-specific terminology, especially in science and history. This aligns with the high frequency of short chat sessions, suggesting that many interactions were brief and focused on content understanding.

Other frequently used tools included “Extract Key Points” (60 uses), “Conversation across time” (58 uses), and “Summarizer” (50 uses). These tools were particularly popular in Ireland, pointing to a more diverse and in-depth use of SB’s learning support features in that context. Additionally, “Debate Tool” (44 uses) and “Text Adaptation” (28 uses) were almost exclusively used by Irish students, highlighting the integration of SB in language-based and reflective tasks such as structured argumentation and creative writing. In contrast, Greece recorded significantly higher usage of the “Conversation Across Time” tool (32 out of 58 total uses), indicating its alignment with History instruction in that setting.

Tools like “Chronological Order of Events” and “Self-Assessment Quiz” were also used across countries, likely supporting curriculum-related tasks such as timeline construction and test preparation. Less frequently used tools included “Revision Tool”, “Interactive Concept Exploration”, and “Auto-Grade Exercise”, suggesting either limited integration into classroom tasks or unfamiliarity with their functionality.

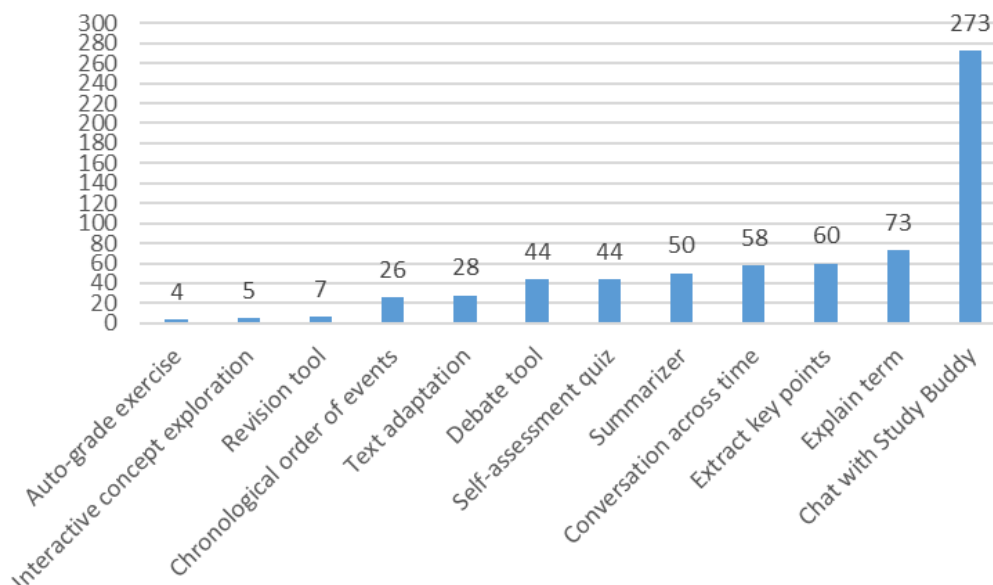


Figure 4.14: Study Buddy Tool Usage

The analysis also revealed notable variation in the diversity of tools used and topics explored across participating countries. Irish students demonstrated the broadest engagement in terms of tool variability, utilizing eleven of the twelve tools available within SB. This suggests that Irish implementations effectively integrated the full range of SB tools and functionalities into a wide array of learning activities.

Thematic variety was equally rich, with Irish students engaging in chats covering local and global history, environmental topics, and structured argumentation, demonstrating how SB supported both curriculum-linked learning and broader critical thinking skills. The SB tools' sustained, systematic, in depth and flexible use across several subjects reflects strong motivation among students, as well as high degree of alignment with varied pedagogical goals. These offer a reasonable explanation for the positive impact of SB in student learning gains, engagement, autonomy, and motivation, as reported in the Irish pilot results.

Greek students engaged with nine different tools within SB, with notably high usage of features aligned to their History curriculum, as illustrated above. Notably, 80 out of 131 total chats initiated by Greek students were related to History topics, indicating that SB was deeply embedded in the experimental group's History instruction. This targeted use helps explain the significant learning gains observed among students in the experimental group compared to those in the control group (section 4.1.5).

Students in Cyprus engaged with seven different tools, primarily utilizing features such as "Explain Term" and "Self-Assessment Quiz" to support conceptual understanding and assessment preparation in Physics and Chemistry. Notably, 51 out of 128 total chats focused on the topic of Gravitational Force in Physics. For this topic, all seven tools were actively used, reflecting a particularly rich and sustained engagement. These findings suggest that the experimental group studying Gravitational Force demonstrated the most in-depth and consistent use of Study Buddy among Cypriot classes, offering a plausible explanation for the statistically significant learning gains observed in this group compared to the control group (section 4.1.2).

Swedish students exhibited limited diversity, using two tools, "Explain Term" and the general "Chat with Study Buddy" feature. This narrower engagement reflects the pilot's limited scope in Sweden as well as the need for further instructional integration to promote broader use.

Overall, the observed usage patterns suggest that students tended to favor features providing immediate academic support, while more specialized, interactive, or advanced tools were used selectively, often depending on the level of teacher scaffolding, student engagement, and alignment with curricular goals. The strong reliance on the chat-based interface highlights the accessibility and user-friendliness of conversational AI for educational contexts. However, usage data from specific experimental groups in Cyprus and Greece, as well as all groups in Ireland, indicate that when SB is systematically embedded into teaching and learning practices, supported by motivated and well-prepared teachers, students are more likely to explore a broader range of tools and engage with diverse, cognitively demanding content. This deeper engagement was associated with notable gains in academic performance, motivation, and overall participation.

These insights suggest that while SB offers a versatile suite of learning tools, its impact is closely tied to how its features are introduced and supported within classroom instruction. Future implementations may benefit from encouraging teachers to promote the use of SB tools through integrated learning activities, enhancing student familiarity and diversifying educational use cases.

4.3.2. Teacher Mate Log File Analysis

This subsection focuses on teachers' interactions with TM during the second pilot cycle. Drawing on log file data, the analysis explores how educators across the four participating countries engaged with TM to support their instructional practices. Specifically, it examines the number of unique teacher users, frequency of use, duration of sessions, and the popularity of different tools. These metrics offer insight into how TM was adopted as a support system within real classroom environments, as well as the extent to which teachers explored its functionalities in practice.

A total of 13 teachers used Teacher Mate during the pilot implementation period. As shown in Figure 4.15, Cyprus had the highest number of active users, followed by Sweden, Ireland, and Greece. While the number of users per country was relatively small, these interactions provide valuable insight into early-stage engagement with AI-assisted teaching support.

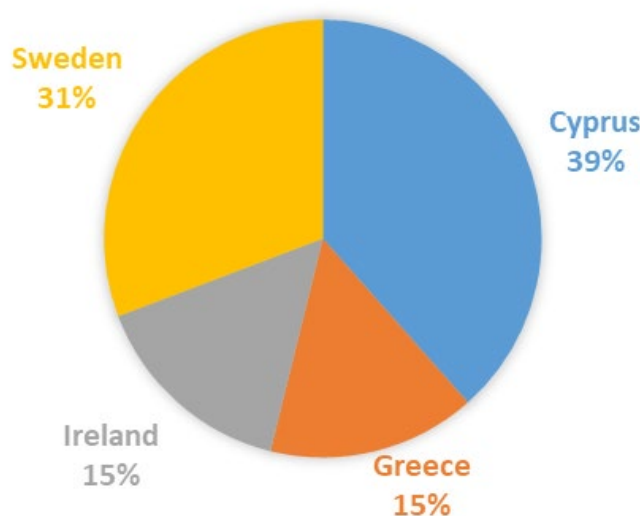


Figure 4.15: Distribution of Unique Teacher Users per Country

Teacher interaction frequency with TM varied considerably. As shown in Figure 4.16, usage ranged from a single session to as many as 10 interactions per teacher. Most teachers used the application between three and five times, suggesting targeted use for specific tasks rather than continuous engagement. Cyprus presented the highest frequency, with one teacher reaching up to ten separate interactions. Ireland and Sweden reported more modest but repeated usage patterns, while no usage frequency was recorded in Greece due to missing log entries.

Session duration was also examined to shed light on the intensity and nature of use. Sessions were categorized into short (0–10 minutes), medium (10–60 minutes), and long (over 60 minutes). As illustrated in Figure 4.17, short and medium sessions were equally common, each representing 22 out of the 53 recorded sessions (42%). Long-duration sessions were less frequent but still notable, comprising 9 sessions (17%). Notably, Ireland and Sweden had the highest number of medium and long-duration sessions, indicating more sustained interaction with the application’s tools. Cyprus showed more frequent but shorter sessions, suggesting quick and focused use cases.

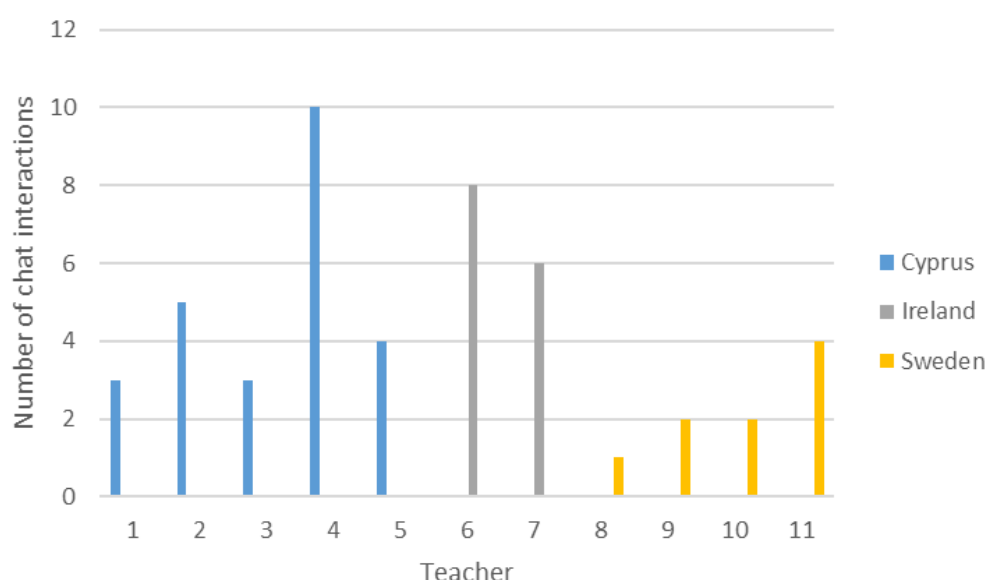


Figure 4.16: Number of individual teacher Interactions

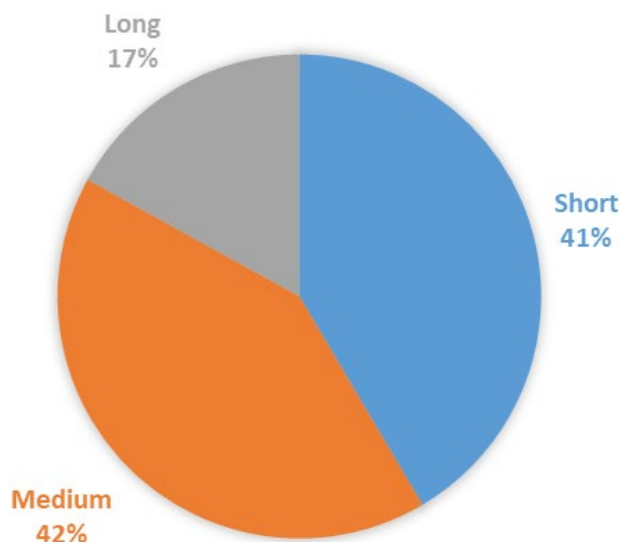


Figure 4.17: Distribution of Teacher Mate Interactions by Duration

Tool usage logs provide valuable insights into how teachers integrated TM into their instructional practice. As shown in Figure 4.18, the most frequently used tool was the Quiz Creator, which was accessed 17 times across pilot sites. This was followed by the “Chat with Teacher Mate” feature (11 uses), the Lesson Plan Generator (5 uses), and the Concept Exploration Assistant (4 uses). These tools reflect key instructional tasks such as assessment design, lesson planning, and content support.

In Cyprus, teachers predominantly used the Quiz Creator, likely to develop quick formative assessments or support classroom review activities. Ireland demonstrated a wider range of tool use, including the Lesson Plan Generator, Presentation Planner, and Text Adaptation, which were likely employed to scaffold lessons and differentiate instruction for diverse learners. Sweden, though showing fewer tool interactions overall, registered several medium and long sessions, suggesting deeper exploratory use of fewer tools. Tools such as Generate Test, Create Rubric, Teaching Material Creator, and Chronological Order of Events saw no usage, which may reflect either a lack of alignment with immediate teacher needs or limited familiarity with these features.

The data suggest that teachers used TM selectively, focusing on tools that directly supported their instructional tasks and time-sensitive planning needs. The frequent use of the Quiz Creator and Lesson Plan Generator indicates demand for automated content creation and classroom organisation. Moreover, the equal distribution of short and medium sessions implies that TM was used both for quick generation of resources and for more extended planning sessions.

The variation in tool usage across countries may be attributed to contextual factors such as curriculum focus, digital literacy levels, classroom demands, and the degree of teacher autonomy during the pilot. While total interactions were modest, the diversity in tool selection and session length signals genuine exploration of TM’s capabilities.

These findings highlight the need for continued professional development to deepen teachers’ familiarity with the full range of TM tools and support the integration of AI tools into everyday pedagogical practices. Structured guidance, training modules, or exemplar use

cases could help expand usage beyond the most immediately intuitive tools and foster more consistent engagement across contexts.

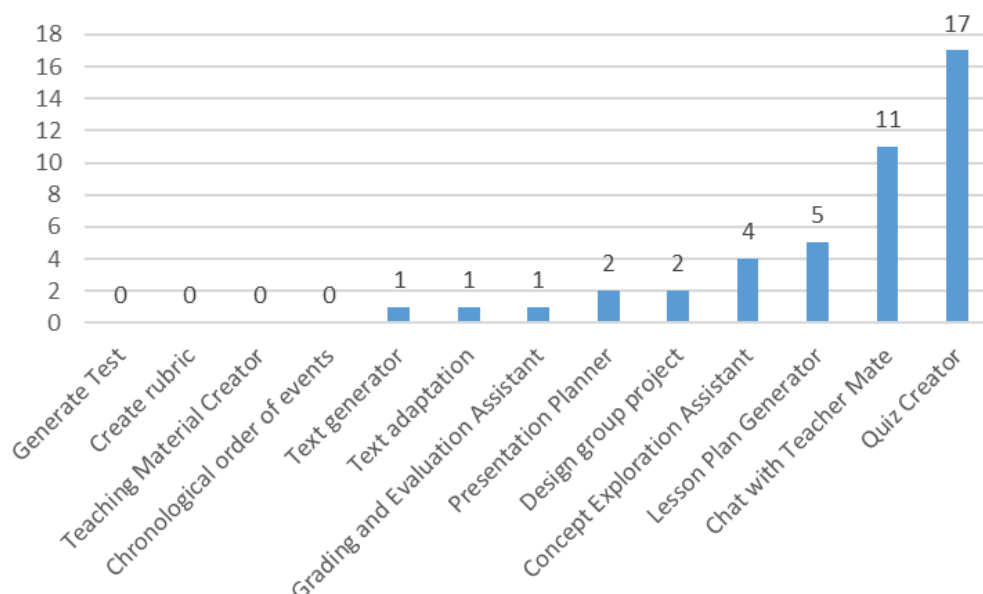


Figure 4.18: Teacher Mate Tool Usage

4.4. Teacher Mate's Influence on Teacher Instructional Practice

To better understand the role of TM in supporting teachers' daily practice, feedback was collected through pre- and post-implementation interviews and surveys with participating educators across all pilot sites. These instruments explored how TM influenced lesson planning, instructional delivery, assessment practices, and classroom dynamics. Teachers also reflected on the extent to which the application aligned with their pedagogical goals, supported personalized instruction, reduced their workload and impacted their professional routines.

4.4.1. Teachers' Initial Practices, and Views on AI Use in the Classroom

Teachers who participated in the AI4EDU pilot entered the program with diverse backgrounds in terms of teaching practices, prior exposure to AI tools, and their expectations regarding the integration of AI applications in education. The responses collected through 16 pre-implementation surveys and interviews provide a comprehensive picture of their initial pedagogical approaches, their familiarity with AI, and their attitudes toward the use of AI technologies.

- **Initial teaching practices**

The majority of teachers reported using a wide range of traditional and digital tools in their teaching. Lesson planning and assessment were largely manual or semi-digital processes, often relying on textbooks, worksheets, PowerPoint presentations, and, in some cases, national platforms. Several teachers mentioned using digital tools like Kahoot, online quizzes, or rubrics in platforms such as Teams. However, many noted a lack of sufficient time for individualized planning and resource creation, which led to a need for support in areas like differentiation and assessment design.

While some teachers reported organizing content collaboratively or drawing on existing material, others emphasized the effort and time required to prepare content that meets the learning needs of diverse students. A recurring theme was the desire for more efficient and targeted lesson planning and the ability to personalize instruction without increasing workload.

- **Personal use and familiarity with AI**

Teachers' prior experience with AI varied widely. Some had limited or no exposure to AI technologies, while others had actively used tools like ChatGPT, Copilot, Notion AI, and personal assistants like Siri or Google Assistant for content creation, personal organization, or creative purposes. A few participants expressed concerns or uncertainty about how AI systems work and whether they could be trusted in educational contexts. Others felt relatively confident, highlighting the potential of AI to assist with tasks like summarizing texts, generating visual material, or designing content.

Despite varying levels of experience, many teachers demonstrated openness to learning and a willingness to explore the potential of AI in supporting their teaching practices. Some expressed the importance of using AI critically and selectively, while a few also mentioned ethical or practical concerns, such as data privacy or AI reliability.

- **Expectations and anticipated benefits of Teacher Mate**

Teachers were generally optimistic about the potential of AI to enhance their work. Their primary expectations included saving time in lesson planning and assessment creation, receiving new ideas and inspiration for content development, and achieving greater alignment with curriculum standards. Many anticipated that AI would help them better differentiate instruction and improve the quality of the materials they used in class.

Several educators expected TM to be a complementary tool that could support their teaching practices without replacing their pedagogical autonomy. Others highlighted that if implemented thoughtfully, it could improve efficiency and free up time for more meaningful classroom engagement. Commonly mentioned anticipated benefits included better organization, access to diverse teaching strategies, and enhanced responsiveness to student needs.

- **Concerns and ethical reflections**

While enthusiasm for AI tools was evident, several concerns were raised. Some teachers questioned the accuracy and reliability of the content generated by AI. There were worries about potential over-reliance on AI, both by teachers and students, leading to diminished critical thinking and creativity. Teachers stressed the importance of maintaining control over instructional decisions and ensuring that AI tools serve as aids rather than replacements.

Others noted the need for thoughtful implementation, highlighting challenges such as integrating AI into existing teaching workflows, data protection, and balancing AI use with ethical and pedagogical standards. Concerns about student dependency on AI, especially for tasks requiring higher-order thinking, were also mentioned. A few teachers advocated for gradual adoption, starting with non-critical tasks and monitoring the outcomes before deeper integration.

- **Expectations for students' use of Study Buddy**

Teachers envisioned SB as a personalized learning assistant that could support students beyond the classroom, particularly in understanding difficult concepts, reviewing content,

and enhancing their autonomy. Many saw it as a tool that could offer tailored support based on students' needs, helping them with self-assessment, content exploration, and clarification of doubts.

At the same time, some educators expressed concern about students using SB passively or as a shortcut to avoid deeper engagement. They emphasized the need to guide students in developing critical digital literacy and to ensure that AI use promotes and not undermines learning. A few teachers suggested that SB should complement class activities and that students must be taught when and how to use such tools meaningfully.

- **Ideas and recommendations for AI Integration**

Participants offered a variety of suggestions for enhancing AI integration in education. These included training programs for both teachers and students, clear usage guidelines, development of pedagogically sound content, and alignment with national curricula. Several educators emphasized the importance of preserving critical thinking, creativity, and human interaction, viewing AI as a support system rather than a replacement for human judgment.

Concerns related to ethics, data privacy, the potential misuse of AI, and the quality of generated content were echoed throughout the responses. Overall, teachers advocated for a balanced approach that combines the efficiency and support offered by AI tools with thoughtful implementation, ongoing reflection, and a strong pedagogical foundation.

4.4.2. Teachers' Reflections on the Use of Teacher Mate

Following the implementation phase of the AI4EDU applications, participating teachers shared detailed reflections on their experiences with TM. Their responses provide insights into the practical integration, perceived benefits and limitations, and future potential of AI tools in real classroom settings.

- **Integration and pedagogical alignment**

Teachers integrated TM into their teaching in various ways, ranging from the creation of quizzes and evaluation tests to the exploration of concepts and development of personalized materials. In some cases, the application was used during live instruction with students, while others used it primarily for lesson planning and content generation. Despite varying levels of integration, often influenced by time constraints or competing school activities, many educators indicated that the application complemented their existing pedagogical approaches and facilitated more efficient lesson preparation. As one teacher noted, *"Using Teacher Mate made planning so much quicker and more effective."*

Several teachers, particularly those with more time and familiarity with the application, noted that TM helped them introduce differentiated instruction, offer extra support to students, and extend learning beyond classroom hours through home assignments. One teacher reflected, *"I reinforced the learning process beyond school hours with quizzes and assignments, promoting deeper understanding and long-term knowledge retention."* However, a few educators reported limited alignment with their current methods, citing technical barriers or uncertainty regarding the accuracy of AI-generated content. As one explained, *"There was some uncertainty about the accuracy of the information it provided."*

- **Perceived benefits and workload impact**

A notable benefit highlighted by teachers was the reduction in time and effort required for planning and resource development. Features such as automated quiz creation, chronological ordering of events, and on-demand explanations were found particularly useful.

One teacher mentioned that TM allowed for “faster development of tests and assessment questions,” while another pointed out that it “*made it easier to create and adapt teaching materials.*”.

Nevertheless, not all teachers experienced a significant reduction in workload. Some struggled with initial usability issues, such as managing student logins or navigating the test creation process, due to limited exposure. As one teacher shared, “*I had problems with student accounts because I didn’t fully understand how they worked, and I had trouble sharing my material. It got better once I realized I had to click on the test icon on the homepage instead of creating a test from scratch.*” A few participants also pointed out that while AI could accelerate certain tasks, it still required close oversight and verification, especially when used for creating instructional content.

- **Challenges and ethical considerations**

Technical challenges were frequently reported, including network issues and login difficulties. Teachers also noted the absence of features such as instant feedback and automated scoring, which would have enhanced self-assessment. One teacher remarked, “*It doesn’t provide answers to the quizzes students complete, so it doesn’t help them evaluate themselves.*”. It should be noted that although these features were already available in the application, some teachers were unaware of their functionality, which confirms their limited exposure to and familiarity with the tool.

Ethical concerns centered on data privacy, overdependence on AI, and the potential erosion of critical thinking. Some expressed unease about students using AI without questioning the validity of its responses, stressing the importance of teaching responsible and reflective AI use. One teacher stressed, “*Students need time and guidance to use it effectively... always with critical thinking and evaluation.*” Several teachers underscored the irreplaceable role of human interaction and emphasized the need for teacher facilitation alongside AI tools.

- **Suggestions for future use**

Teachers proposed several recommendations for enhancing the integration of AI applications like TM and SB into broader educational practices. A recurring theme was the need for a unified and more intuitive user interface. One Swedish teacher suggested that “*all tools would function better if integrated into a single chat-based interface instead of being separate applications.*”.

There was also a strong call for comprehensive training for both teachers and students, not only on how to operate the various tools but also on how to use them pedagogically and ethically. As one teacher put it, “*Teachers and students need proper training to use the tools critically and effectively.*”.

In addition, teachers emphasized the importance of incorporating curriculum-aligned content, age-appropriate scaffolding, and customizable options that fit specific subject areas and learner profiles. Some advocated for building the tools on open and flexible platforms to ensure long-term adaptability and sustainability. As one participant recommended, “*If the school is going to invest in an AI solution, I suggest it be a long-term, sustainable platform built on open infrastructure.*” Others emphasized the importance of self-assessment features, suggesting the addition of “*tools that help students evaluate themselves and track their progress more independently.*”.

- **Continued use and outlook**

The majority of teachers expressed a positive attitude toward continued use of the applications, especially if they are further developed and supported at the institutional level. While some educators remained neutral or cautious, citing the need for technical refinements and better alignment with their teaching goals, most recognized the potential of AI tools to support more personalized, efficient, and engaging learning experiences. One teacher summarized this outlook by stating, *“I can see how AI tools could support student engagement by providing tailored feedback. We need to keep exploring new methods.”*.

5. Synthesis of Findings

This chapter synthesizes the main findings from the analysis of data collected during the second pilot cycle of the AI4EDU applications, SB and TM. Drawing on quantitative and qualitative evidence presented in Chapter 4, this section revisits the research questions outlined in Chapter 2 and offers a consolidated interpretation of the educational impact of the applications. The aim is to assess the extent to which the applications contributed to enhancing teaching and learning practices across the four participating countries. Responses to the research questions are based on the triangulation of results from academic tests, motivation and engagement surveys, application usage logs, teacher interviews, student feedback, and classroom observations.

5.1. Student-Related Research Questions

RQ1: How does SB support students in preparing for assessments and tracking their academic progress?

Findings from log data, survey responses, and student interviews suggest that SB played a modest but meaningful role in supporting students' preparation for assessments. Students reported using the application to generate flashcards, complete self-assessment quizzes, and summarize content in preparation for tests and assignments. These functionalities helped them organize material and engage in focused revision. In several cases, teachers incorporated SB into homework and project activities, further scaffolding its use for exam readiness. Although the application includes features for performance tracking and feedback, these were largely underutilized during the pilot due to time constraints and limited familiarity with the applications. As a result, SB functioned more as a short-term revision aid than as a tool for systematically monitoring academic progress over time.

RQ2: To what extent does SB improve students' understanding and retention of key concepts?

Case studies across pilot sites suggest that SB contributed meaningfully to students' understanding and retention of key concepts, particularly when its use was integrated with teacher guidance and curriculum-aligned activities, as reflected in gain scores and test outcomes. In Greece and Cyprus, students in experimental groups generally outperformed control groups in post-tests, with the History case in Greece showing a statistically significant improvement. These results imply that SB can facilitate content comprehension, especially when used to reinforce classroom instruction.

Qualitative data from Ireland and Sweden further illuminate how SB supported conceptual understanding. Students consistently reported that features like “Explain Term,” “Summarizer,” and the open-ended “Chat” function helped them clarify difficult topics, rephrase complex content, and engage in self-directed exploration of subject matter. These tools were especially valued for subjects requiring definition-based or narrative comprehension, such as History and Biology.

In Ireland, where SB was used in classroom settings, teachers observed that students relied on it to revisit and rephrase explanations in their own words, fostering active cognitive engagement. Some students also described using SB to check their reasoning or compare answers to test questions, thus reinforcing retention through self-verification.

Teachers across countries noted that SB's ability to present information in digestible, stepwise formats helped reduce cognitive overload, allowing students to build conceptual understanding gradually. The immediacy of response and the possibility to ask follow-up questions also encouraged iterative learning, which is essential for deeper processing and long-term retention.

Despite these positive effects, the impact of SB on conceptual learning varied depending on how systematically it was integrated into classroom routines. In cases where students used SB independently without clear scaffolding or goals, its influence appeared more limited. This underscores the importance of teacher-facilitated use and alignment with pedagogical intentions.

RQ3: How does the use of SB impact student engagement and motivation in the learning process?

The impact of SB on student engagement and motivation was multi-faceted, with modest trends observed across both quantitative and qualitative data. Results from the MES showed generally stable scores pre- to post-implementation, with no statistically significant differences between experimental and control groups. Nevertheless, small positive shifts were consistently observed in the experimental groups. These trends, while not conclusive, suggest that SB may contribute to maintaining or slightly enhancing students' motivational profiles when used appropriately.

Qualitative data provided a deeper understanding of SB's motivational influence. In Ireland, where students used SB regularly during classroom sessions, many reported feeling more engaged and confident in navigating complex topics. Students appreciated the conversational format, interactive feedback, and autonomy afforded by SB's tools. This sense of agency was particularly important in helping students overcome reluctance to participate in traditional classroom discussions. One student noted that *"it's easier to ask questions to the app than in front of everyone"*, highlighting how SB helped lower social barriers to engagement.

Teacher reports and classroom observations further confirmed these effects. In inclusive classrooms SB appeared to support sustained attention and increased participation, particularly among students with SEN or limited language proficiency. The ability to receive immediate, individualized support fostered a sense of control and reduced frustration, which in turn contributed to higher emotional engagement.

Students also described using SB as a motivational resource during independent study, appreciating its capacity to explain topics at their own pace. Features such as flashcard generation, summarization, and targeted concept explanation were frequently mentioned as making learning more dynamic and less overwhelming.

However, the motivational impact of SB was strongly linked to how it was introduced and supported by teachers. In cases where implementation lacked structure or purpose, students tended to engage with the application superficially or abandon it altogether. This indicates that while SB has potential to foster motivation and engagement, its success depends heavily on pedagogical framing, consistent use, and alignment with classroom goals.

RQ4: How do students perceive the usability and effectiveness of SB in enhancing their learning experience?

Students' perceptions of SB's usability and effectiveness were mixed but pointed to several areas of promise. Quantitative data from student questionnaires showed that a substantial number of students remained neutral in their responses, particularly regarding overall satisfaction and ease of use. However, a notable proportion of students expressed positive views about SB's clarity, interactivity, and usefulness in test preparation and content review. These perceptions were especially strong in contexts where SB was systematically embedded into classroom routines and supported by teacher guidance.

Open-ended responses and interview data from Ireland offered more nuanced insights. Students described the application as *"easy to use once you know where everything is"* and praised features like "Explain Term" "Summarizer" and the "Chat" function for helping them grasp difficult concepts. The interactive nature of the conversational interface was seen as more engaging than static resources, with some students highlighting the benefit of being able to *"ask questions at any time without feeling embarrassed."*

Despite these positive experiences, several usability issues were reported. Some students found the interface confusing at first, especially when navigating between tools or locating past conversations. Others noted occasional inaccuracies in content generation, which undermined their trust in the application. A few students mentioned the lack of multimedia support and felt that responses were "too text-heavy" or "not always easy to follow."

Perceptions of SB's usefulness were closely tied to how the application was introduced. In classrooms where teachers modeled its use, assigned specific tasks, integrated it in lesson plans or aligned it with curricular objectives, students reported greater satisfaction and learning benefits. Conversely, when students were left to explore the application independently, some struggled to see its relevance or stopped using it altogether.

These findings suggest that while SB offers several features students find valuable, its usability and perceived effectiveness depend on clear onboarding, consistent use, and pedagogical integration. Addressing interface clarity, improving content accuracy, and offering scaffolded support could enhance students' experience and increase adoption across diverse learner groups.

RQ5: What learning patterns and trends can be observed from students' interactions with SB, and how do these correlate with their academic outcomes?

Log file analysis revealed varied usage patterns among students, with most engaging in short, focused interactions and a smaller number demonstrating more frequent or extended use. The majority of students used the application between one and five times, typically in sessions under 10 minutes. These interactions were often centered around immediate needs, such as clarifying concepts or preparing for assessments, suggesting that SB functioned primarily as an on-demand support tool.

More sustained usage was observed in specific contexts, particularly in Ireland, where SB was embedded into classroom instruction and guided activities. In such cases, students exploited the full potential of SB tools, focusing on "Chat with SB" "Explain Term" and "Summarizer" and often revisiting content multiple times. This structured integration appeared to encourage deeper engagement with the application's educational functionalities.

Usage patterns also reflected the pedagogical approach adopted by teachers. Where teachers explicitly integrated SB into classroom activities, homework, or revision assignments, students demonstrated more consistent engagement. In contrast, more sporadic and superficial use was observed in contexts where SB was introduced without structured guidance.

Overall, the analysis suggests that SB's educational value is shaped by how it is introduced and supported within the classroom context. Structured implementation and teacher scaffolding were associated with more frequent and meaningful use.

5.2. Teacher-Related Research Questions

RQ1: How does TM affect teachers' efficiency in lesson planning, instructional delivery, and assessment tasks?

Findings from log data and teacher feedback indicate that TM contributed to improving teachers' efficiency primarily in lesson planning and assessment preparation. Tools such as the Quiz Creator and Lesson Plan Generator were among the most frequently used, enabling teachers to quickly generate tests, assignments, and structured lesson outlines. Teachers in Cyprus and Ireland, in particular, reported time savings and greater ease in organizing content and structuring instructional sequences, highlighting its practical value in everyday routines.

Despite these benefits, some teachers experienced initial difficulties navigating the interface or locating specific features, which limited early use. These issues were often resolved over time, suggesting that TM's efficiency benefits increased with familiarity. Instructional delivery was less directly affected, as TM was mostly used as a planning tool outside classroom hours. Still, by reducing preparation time and providing ready-made materials, TM indirectly supported smoother delivery of lessons. Overall, teachers who actively engaged with TM reported improved workflow and appreciated the structured support it offered for core teaching tasks.

RQ2: In what ways does TM support teachers in providing personalized instruction to address diverse student needs?

Teachers reported that TM supported personalized instruction primarily through its ability to generate adaptable materials tailored to different student profiles. In Ireland, for example, educators used the Text Adaptation tool to simplify content for students with SEN and EAL learners as well as the Quiz generation tool to create individualized quizzes. This allowed them to provide differentiated materials that aligned with students' varying language proficiency and cognitive levels.

The Concept Exploration and Lesson Plan Generator tools were also used to adjust instructional content, helping teachers introduce or reinforce key concepts at different levels of complexity. Teachers appreciated the ability to modify inputs and prompts to generate materials that suited specific classroom contexts or individual learners. However, while the tools offered flexibility, the extent of personalization was largely teacher-driven. As such, TM functioned as a flexible resource generator rather than a fully personalized instructional system.

RQ3: How effective is TM in helping teachers monitor student progress and identify areas for improvement?

Teachers found TM useful for preparing assessments, such as quizzes and tests, which could indirectly support monitoring student understanding. However, its effectiveness as a tool for

tracking student progress was limited during the pilot. While the application includes performance tracking features, these were rarely used, mainly due to time constraints, limited training, and the lack of integration with classroom routines.

Educators highlighted the need for more illustrative visual summaries of student results as a barrier to using TM for systematic progress monitoring. TM's potential to support data-informed teaching remained underutilized in practice, pointing to the need for more training and exposure to the tool as well for improvement of feedback mechanisms to facilitate its use in this area.

RQ4: What is the perceived impact of TM on reducing teacher workload?

Teachers generally viewed TM as a valuable aid in reducing their workload, particularly in time-consuming tasks such as lesson planning, quiz creation, and resource development. Tools like the Quiz Creator and Lesson Plan Generator were especially appreciated for enabling quick content generation, reducing the time and effort needed to prepare personalized and structured materials. In some cases, teachers reported that TM helped them manage planning more efficiently and inspired new instructional ideas, contributing to both time savings and improved lesson quality.

Despite these benefits, the degree of workload reduction varied. Teachers who were already comfortable with digital tools and had sufficient time to explore TM experienced more noticeable advantages. In contrast, those who encountered usability challenges, such as difficulty navigating the interface, managing student accounts, or sharing materials, felt that the initial learning curve offset some of the expected time savings.

Overall, TM was perceived as a supportive application with clear potential to streamline teaching-related tasks, particularly when supported by adequate training and user guidance.

RQ5: How do teachers evaluate the overall impact of TM on student learning outcomes?

Teachers generally believed that TM indirectly contributed to improved student learning by enhancing instructional quality and enabling more personalized content. By simplifying the creation of quizzes, lesson plans, and differentiated content, TM enabled more organized and responsive teaching, which teachers believed positively influenced student understanding and engagement. Several educators noted that students were more focused and better prepared when lessons were supported by TM-generated materials.

However, teachers also emphasized that TM's contribution to learning outcomes depended heavily on how it was integrated into instruction. They stressed that the tool was most effective when used as a complement to established pedagogical practices, not as a standalone solution. Some teachers expressed uncertainty about the accuracy of AI-generated content, which occasionally limited their confidence in using it extensively for student-facing materials.

Overall, while teachers did not directly attribute learning gains to TM, they recognized its value in enhancing their instructional practice in ways that could benefit students and support differentiated instruction, particularly in terms of organization, differentiation, and instructional clarity.

This synthesis confirms that both SB and TM have educational value when integrated thoughtfully into teaching and learning. While the short duration and uneven implementation of the pilots constrained the magnitude of measurable outcomes, the findings highlight promising patterns, especially in inclusive and well-supported settings. These insights can

inform future scaling, teacher training, and the design of AI applications that are responsive to the realities of classroom practice.

6. Conclusions and Recommendations

This chapter presents the overarching conclusions and strategic recommendations emerging from the AI4EDU project's second pilot cycle. Building on comprehensive evidence gathered through quantitative assessments, qualitative reflections, and application usage data across four European countries, the analysis confirms that AI-powered applications such as SB and TM hold considerable potential to enhance teaching and learning, provided they are integrated thoughtfully, systematically, and with strong institutional support.

6.1. Conclusions

Pedagogical integration

The pilot implementations of SB and TM demonstrated that AI tools can support improved teaching and learning, especially when pedagogically embedded in structured classroom contexts. Quantitative data alone showed modest or inconclusive trends, but triangulated with qualitative evidence, the findings suggest that SB supported student motivation, engagement, and academic understanding, particularly in settings where AI use was teacher-guided, curriculum-aligned, and embedded in classroom practice. The Irish implementation notably exemplified how structured integration, teacher facilitation, and institutional support can drive student agency, deeper engagement, and critical thinking.

Teacher engagement

Teacher readiness, motivation, and sense of ownership emerged as central enablers for successful AI adoption. Classrooms where teachers were confident, well-informed, and actively engaged in co-designing AI integration saw higher levels of student interaction with the tools and more meaningful learning outcomes. In contrast, contexts where AI was introduced as an auxiliary homework tool, with minimal classroom scaffolding, often led to superficial engagement and underutilization of features. These patterns underscore the importance of professional development, peer collaboration, and trust in AI tools to foster effective implementation.

Inclusive potential of AI-supported learning

SB showed encouraging results in promoting inclusivity and personalized support for students with SEN, non-native language speakers, and students from marginalised backgrounds. In the Irish DEIS school context, AI-supported classrooms showed increased student engagement, language development, and emotional confidence, particularly among Ukrainian and Roma students. Adaptive feedback, non-judgmental responses, and opportunities for self-paced learning contributed to more equitable classroom dynamics. These findings highlight the importance of designing AI applications with inclusive features and ensuring teacher capacity to harness them in differentiated instruction.

Organisational readiness and institutional culture

Successful integration of AI tools did not depend solely on classroom-level decisions. Organisational enablers, such as collaborative school cultures, psychological safety, distributed leadership, and aligned digital strategies, amplified the educational value of AI tools. Schools with shared leadership and a culture of innovation demonstrated greater agility and collective learning in adapting AI tools to local needs. Conversely, fragmented leadership,

limited coordination, or lack of higher-level vision constrained deeper integration. These systemic elements point to the need for institutional capacity-building and policy frameworks that go beyond individual teacher readiness.

Challenges in data interpretation and tool adoption

The short duration of the pilot, along with variability in teacher engagement and classroom conditions, limited the ability to detect statistically significant changes through quantitative metrics alone. Student self-report surveys, while useful, did not always align with observed behaviours or learning gains. Moreover, several SB and TM features remained underused due to insufficient visibility, lack of integration into lesson plans, or limited onboarding. These challenges indicate that the success of AI in education depends not only on tool design but also on clear guidance, structured implementation, and supportive ecosystems.

6.2. Recommendations for the Educational Community

The experiences and findings from the AI4EDU pilot implementations reveal that the successful integration of AI tools like SB and TM hinges on more than just technological availability, as it requires a transformation in teaching culture, practice, and support structures. Central to this transformation is the role of educators, whose confidence, engagement, and pedagogical intent shape the effectiveness of AI in the classroom.

To begin with, professional development must be comprehensive and sustained. Teachers need more than technical know-how; they require practical, context-sensitive guidance on embedding AI into their instructional practices. This includes training on how to align AI tools with lesson objectives, how to scaffold student use of AI in inquiry-based and reflective learning, and how to critically assess AI-generated outputs. The pilot findings suggest that early exposure and guided experimentation with SB and TM increased teacher confidence and willingness to innovate, leading to richer and more sustained classroom use.

Additionally, teachers should be afforded flexibility and autonomy in adapting AI tools to suit the diverse needs of their students. When educators have the freedom to tailor AI-supported activities, such as adapting reading levels, generating formative assessments, or facilitating differentiated feedback, they are better able to meet students where they are. This adaptability is particularly crucial for reaching learners with SEN, non-native language speakers, and students from marginalized backgrounds.

Institutional support is also essential. Peer collaboration and shared learning among teachers foster communities of practice where implementation strategies can be refined, resources shared, and barriers addressed collectively. In schools where collaborative cultures were encouraged, teachers reported greater motivation, resilience, and innovation in their use of AI tools. These environments also helped to amplify the systemic benefits of AI integration by reinforcing collective ownership and distributed leadership.

Finally, educators must be equipped not only as users of AI tools but as co-designers of their educational application. By actively participating in the iterative refinement of AI-based interventions, through feedback, experimentation, and collaborative reflection, teachers help ensure that these tools evolve in ways that are both pedagogically relevant and ethically grounded. Their input is invaluable in identifying both benefits and limitations that developers and policymakers might overlook.

In sum, the educational community should not treat AI as a standalone solution but as a catalyst for pedagogical innovation. For this potential to be realized, AI tools must be

embedded within a broader ecosystem of teacher empowerment, curricular integration, and collaborative practice.

6.3. Policy Recommendations

The AI4EDU pilot underscores that meaningful AI integration in education requires more than innovation at the classroom level, it demands systemic support through coherent, forward-thinking policy frameworks. These policies must address infrastructure, inclusion, professional development, and ethical implementation in order to maximize the potential of AI tools like SB and TM across diverse educational contexts.

First and foremost, policymakers must prioritize equitable access to digital infrastructure. Reliable internet connectivity, adequate numbers of student devices, and consistent technical support are foundational for AI integration, yet remain uneven across and within European education systems. Addressing these infrastructure gaps is essential to ensure that all schools, not just well-resourced ones, can benefit from AI-enhanced teaching and learning.

Second, inclusion must be at the core of AI design and deployment. The AI4EDU pilots demonstrated that AI tools can enhance learning outcomes for students with SEN, non-native language learners, and students from marginalized backgrounds, when those tools are adaptable and supported by thoughtful implementation. Policymakers should therefore support the development of AI applications that are linguistically and culturally responsive, easily customizable, and capable of supporting differentiated instruction.

Alongside inclusive design, there is an urgent need to invest in teacher capacity-building. Policy must support structured, ongoing professional development that equips educators not just to operate AI tools, but to integrate them meaningfully into curriculum planning, formative assessment, and inclusive pedagogy. National and regional strategies should include training programs co-developed with educators, model implementation cases, and teaching resources aligned with local curricula. Teachers should not be seen as end-users but as co-creators in shaping how AI is used in education.

Equally important is the promotion of school-level readiness and leadership. Findings from the pilots show that distributed leadership, shared decision-making, and a psychologically safe school culture significantly increase the likelihood of successful AI implementation. Policy should therefore encourage leadership models that empower schools to adapt AI tools to their unique contexts. This may involve funding for school-based innovation projects, support for inter-school networks, and mechanisms for collaborative planning between teachers, administrators, and developers.

Furthermore, evaluation and monitoring practices must evolve to reflect the complexity of AI-enhanced learning environments. While quantitative metrics remain important, they must be complemented by methods that capture qualitative dimensions such as student motivation, critical thinking, emotional engagement, and teacher agency. Establishing robust, multi-dimensional evaluation frameworks will ensure that policy decisions are informed by both measurable outcomes and the lived experiences of teachers and students.

Finally, AI policy in education must be guided by ethics and long-term sustainability. Issues such as data privacy, algorithmic transparency, and student autonomy must be proactively addressed. Policymakers at national level should develop clear ethical guidelines and accountability frameworks that protect learners' rights while enabling innovation. Moreover, investments in AI must be directed toward open, flexible platforms that can evolve alongside educational needs, rather than locking schools into proprietary ecosystems.

In summary, the path to effective and ethical AI integration lies in policies that are inclusive, participatory, and systemically grounded. By prioritizing teacher empowerment, infrastructure equity, institutional readiness, and ongoing dialogue between all stakeholders, policymakers can help ensure that AI becomes a tool for meaningful, transformative change in education across Europe.

6.3.1. Final Remarks

The AI4EDU pilot has yielded rich insights into the complex, context-dependent nature of AI integration in education. SB and TM demonstrated promising potential to support differentiated instruction, enhance student motivation, and reduce teacher workload, particularly when used within structured, inclusive, and pedagogically grounded frameworks. However, the presence of the applications alone does not guarantee educational impact. Their effectiveness hinges on thoughtful implementation, robust teacher preparation, and supportive institutional ecosystems.

What emerges most clearly is the central role of educators: their confidence, creativity, and sense of ownership shape whether and how AI is used to its full pedagogical potential. AI tools must therefore be positioned not as replacements, but as allies that augment human teaching, support student autonomy, and expand opportunities for personalised learning. As AI technologies continue to evolve, education systems must strike a balance between innovation and intentionality, ensuring that equity, ethics, and inclusion remain at the heart of implementation.

Moving forward, the lessons from AI4EDU should inform both practice and policy across Europe. With sustained investment in teacher development, inclusive design, and institutional readiness, AI can become a powerful enabler of more engaging, equitable, and effective learning for all.

7. References

Lee, W., & Reeve, J. (2012). Teachers' estimates of their students' motivation and engagement: Being in synch with students. *Educational Psychology*, 32(6), 727–747. <https://doi.org/10.1080/01443410.2012.732385>