EDUBOTS

EDUBOTS Best practices

Prepared by Faculty of Organization and Informatics, in collaboration with the consortium partners in the project

Date of latest revision: 31th January 2022

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Table of contents

Introduction **6**

1. Differ and Hubert updates **8**

1.1. Differ 8

1.2. Hubert 9

2. Pilot 1 courses **13**

2.1. Differ - onboarding educators and students 13

2.2. Hubert - onboarding the educators 16

2.3. Best practices in using chatbots – selected course case studies 19

2.3.1. First case study – Software Engineering course (FOI UniZG) 19

2.3.2. Second case study – Text and Image Processing Course (FOI UniZG) 20

2.3.3. Third case study – Design for all (CYENS/UniCyprus) 23

2.4. Interactions in Differ and Hubert 24

2.4.1. Analysis of system data in Differ 24

2.4.2. Analysis of data from the course evaluations in Hubert 25

2.5. Results of the surveys about Differ 26

2.5.1. Differ and learning process 26

2.5.2. Usability and User Experience of Differ 30

3. Pilot 2 courses **33**

3.1. Best Practices in using Differ and Hubert – selected course case studies 33

3.1.1. Courses at the Cyprus Universities 33

3.1.2. Courses at the University of Granada 38

3.1.3. Courses at the University of Leeds 40

3.1.4. Courses at the University of Zagreb 43

3.2. Interactions in Differ and Hubert 52

3.2.1. Analysis of system data in Differ 52

3.2.2. Analysis of data from the course evaluations in Hubert 53

3.3. Results of the surveys and interviews about the chatbots 53

3.3.1. Differ and learning process 53

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

3.3.2. Differ and teaching process 57
3.3.3. Usability and User Experience of Differ 61
3.3.4. Usability and User Experience of Hubert 64
3.3.5. Results from the interviews with the students and educators 65
4. Pilot 3 courses 69
4.1. Best Practices in using chatbots – selected course case studies 69
4.1.1. Courses at the Cyprus Universities 69
4.1.2. Courses at the University of Granada 71
4.1.3. Courses at the University of Leeds 72
4.1.4. Courses at the University of Zagreb 74
4.2. Interactions in Differ and Hubert 75
4.2.1. Analysis of system data in Differ 75
4.2.2. Analysis of data from the course evaluations in Hubert 75
4.3. Results of the surveys and interviews about the chatbots 76
4.3.1. Differ and learning process 76
4.3.2. Differ and teaching process 79
4.3.3. Usability and User Experience of BO within Differ 85
4.3.4. Usability and User Experience of Hubert 89
5. Pilot 4 courses 93
5.1. Best Practices in using chatbots – selected course case studies 93
5.1.1. Courses at the Cyprus Universities 93
5.1.2. Courses at the University of Granada 95
5.1.3. Courses at the University of Leeds 96
5.1.4. Courses at the University of Zagreb 97
5.2. Interactions in Differ and Hubert 98
5.2.1. Analysis of system data in Differ 98
5.2.2. Analysis of data from the course evaluations in Hubert 99
5.2.3. Hubert meta-evaluation 102
5.3. Results of the surveys and interviews about the chatbots 103
5.3.1. Differ and learning process 103
5.3.2. Differ and teaching process 105
5.3.3. Usability and User Experience of BO in Differ 110

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

5.3.4. Usability of Hubert 112
5.3.5. Results from the interviews with the students and educators 116
6. Best practices in using chatbots – research papers 120
Conclusion 128
References 131

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Consortium

The table presented below shows the consortium of the EDUBOTS project.

Table 1. Consortium partners of the EDUBOTS project

DIFFER	P1 – EDTECH FOUNDRY AS
HUBERT	P2 – ANNA & HUBERT LABS AB
UNIVERSITY OF LEEDS	P3 – UNIVERSITY OF LEEDS
foi	P4 – UNIVERSITY OF ZAGREB, FACULTY OF ORGANIZATION AND INFORMATICS
CYENS	P5 – CYENS CENTRE OF EXCELLENCE
UNIVERSIDAD DE GRANADA	P6 – UNIVERSIDAD DE GRANADA

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Introduction

This document summarizes the best practices which were identified in four pilot rounds of the ERASMUS+ project “Best practices of pedagogical chatbots in higher education / EDUBOTS”. Four partner institutions participated in the pilots which focused on the implementation of chatbots in their teaching practice: University of Granada (Spain), University of Leeds (United Kingdom), University of Zagreb (Croatia), and CYENS Centre of Excellence (formerly known as RISE) from Cyprus which coordinated participation of three public universities of Cyprus - University of Cyprus, Cyprus University of Technology, and Open University of Cyprus.

Practices are based on the courses in which students and educators used educational chatbots Differ and Hubert, while some also used Telegram. Pilots were conducted from January 2020 to December 2021 and each was followed by surveys and interviews.

Differ (https://www.differ.chat/) is a chat platform aimed at users in higher education institutions. It stimulates student collaboration (to connect and chat, find answers, or get help from peers), and supports educators in the communication within the courses (creation of course communities and sharing of additional information with students). Differ offers some functionalities of the chatbots, like pairing up students to facilitate icebreaking and introductions. As Differ is claiming to be a safe place for students to get to know each other and communicate within the context of communities, group chats, and direct messaging, it is vital to ensure full compliance with the European laws such as General Data Protection Regulation (GDPR) and data privacy protection. All servers of Differ are hosted on the Google Cloud within the European Union (Data Center: St. Ghislain, Belgium). The data analytics tool Mixpanel, which is used to analyse user behaviour inside Differ, is also hosted on servers inside the EU (Eemshaven, Netherlands). For further verification, Privacy Policy, Security Statement as well as Terms of Service can be viewed on Differ’s website.

Hubert (https://www.hubert.ai/) is an artificial intelligence (AI) powered chatbot used to automate feedback to educators. It facilitates formative assessment and personalized feedback in the form of structured interviews. Educators set up an evaluation using the provided educational templates with several questions and then share a link with the students who evaluate a course or self-reflect on learned topics by providing narrative feedback.

In addition to Differ and Hubert, a chat platform Telegram (https://telegram.org/) was introduced in Pilot 3 and 4 at the University of Granada. Telegram is a standard messaging platform with web, mobile, and desktop clients for all popular operating systems. Telegram also has an open API, allowing anyone to create chatbots with any degree of intelligence. Additionally, the fact that one can use the platform pseudonymously is an advantage, especially for educators that don’t want their personal phone number known by the students, or students who don’t want it to be known by peers or the educators.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Structure of this document is the following: the next section provides descriptions of Differ and Hubert upgrades throughout the pilots, followed by the sections in which each pilot is described in more detail. Focus on Pilot 1 was on onboarding educators to successfully introduce Differ and/or Hubert into the courses and then collecting information from educators about course content, students, pedagogical approaches and ways of implementing the applications. All case studies are described in a separate document entitled *EDUBOTS Case studies*. Then three course case studies were identified as the best cases in Pilot 1 and presented here. Pilot 1 also served as a starting point for exploring how chatbots were used in the courses: through analysis of interaction data extracted from Differ and Hubert, and analysis of surveys data conducted with the students about their learning process and their perception of usability and user experience with Differ. Pilot 2, 3 and 4 had a more structured approach in implementing and exploring usage of chat apps and chatbots, with a focus on both quantitative and qualitative data collected with the surveys and interviews from students and educators as well. Usability and user experience of Hubert were also explored in the later pilots.

This document presents the results of efforts of all consortium partners to investigate the peculiarities of the chatbots introduced into higher education courses. Both advantages and disadvantages of implemented approaches are presented. Initial results from the surveys and interviews give insights into opinions of educators and students, together with possible directions to further improve applications and implementation approaches. Deeper insight into research results are presented in fourteen research papers which are presented at the conferences or submitted/accepted for publishing in journals, and here are presented with abstracts in the last section, followed by concluding remarks.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

1. Differ and Hubert updates

In this section we present implemented changes in the Differ and Hubert applications which resulted from conducted pilots.

1.1. Differ

Changes in Differ after Pilot 1

The support of chatbots in Pilot 1 was limited to welcoming of students inside a chat (Direct Messaging) as well as some very basic onboarding procedures. While looking at usage patterns within the period of Pilot 1 as well as analyzing results from WP2 Needs Analysis, developers decided to enhance the chatbot experience in terms of an innovative user interface as well as the overall role of the chatbot as a facilitator and match maker. Specifically, a new chatbot area within the app of Differ has been designed, that was focused on discovering possible chatbot conversations for different purposes (situations students/educators find themselves in when starting with a new semester).

List of available chatbot conversation in Pilot 2:

- For Educators: General and educator specific onboarding, Assisted setup of student communities, Preparing group chat enabled FAQs. Split Students into small groups (“break-out chats for discussing specific topics within small groups”).
- For Students: General and student specific onboarding, Match making with other students, Creating question group chats within the context of a community.

Changes in Differ after Pilot 2

Based on the feedback from the Pilot 2, Differ drastically improved the chat and visual experience of the formally introduced chatbot “BO”. Conversational design within the chatbot has been improved as well as the user interface experience such as introducing animated images to illustrate the actions users can take with the chatbot and various control elements to better navigate through the chatbot conversations (Undo/Redo Conversation Steps, Restart, change speed of conversation animation), which could be seen in Figure 1.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

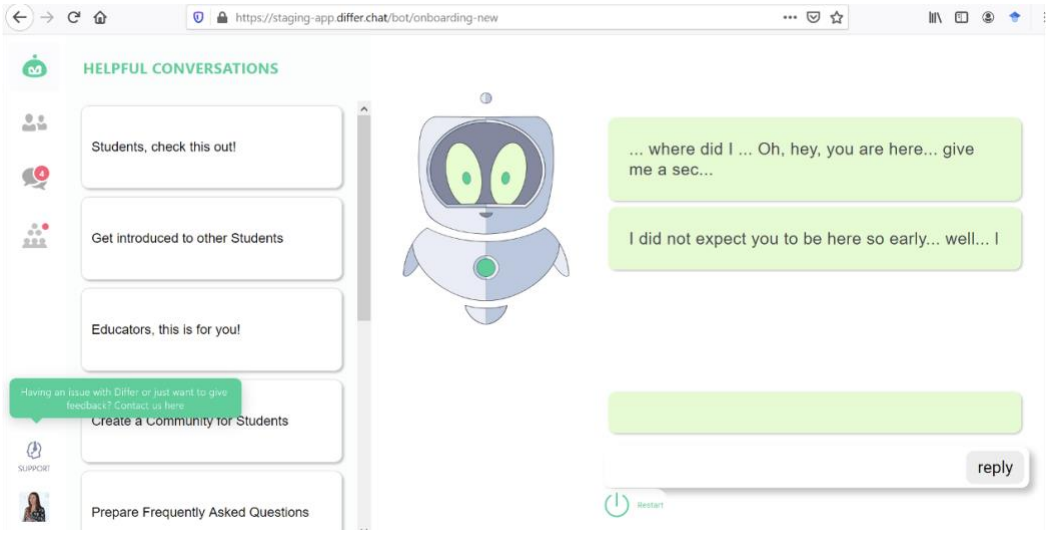


Figure 1. Chatbot BO within Differ

Changes in Differ after Pilot 3

Based on the results of the pilot 3 the focus of further development was on making the chatbots conversations multilingual, supporting languages of all of our partner universities (Norwegian, Greek, Croatian, Spanish in addition to English). Furthermore, an integration of Hubert into the Differ system has been accomplished.

1.2. Hubert

Changes in Hubert after Pilot 1

The first pilot set out as a reference point for coming pilots and used the Hubert version that was commercially available at the time (see Figure 2).

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

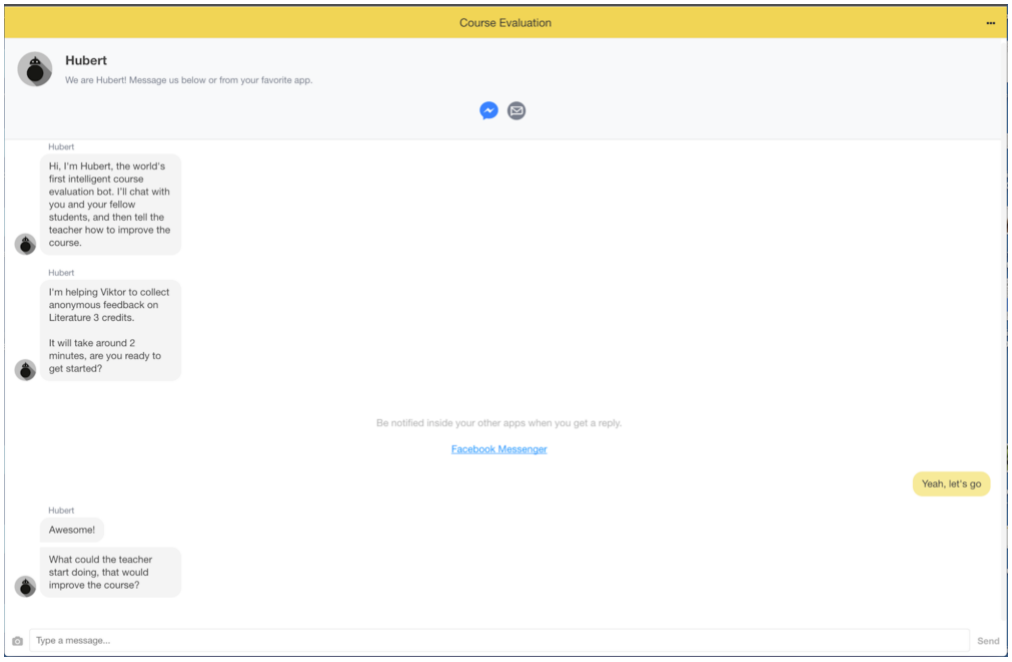


Figure 2. Conversation in Hubert (first interface)

Developers were aware that the trial version had minor bugs and that certain aspects of the user experience were sub-optimal. The initial version of Hubert used in Pilot 1 was enhanced in later pilots by considering user feedback and users' needs reports, which allowed to form an improvement plan that is presented in Table 2.

Theme and needs	**Action taken:**
Bug fixing	
The data export is buggy	Fixed bugs in export
Hubert sometimes misinterprets what students mean	Added more intent training
Export button is not working	Fixed button
Update UX	
Hubert is a bit hard to navigate and find what you need	UX was updated
Adjust the analytics page to become more intuitive	UX was updated
Adjust Hubert's personality to become more of a character	Added more eccentric/fun responses to Hubert's repertoire
Build support for cultural references	Dropped due to time restraints
Languages	
Add more language support	Added language support for Spanish, Greek, and Croatian language

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Integration	
Integration into other systems is needed. Educators want to use minimum systems	Research relevant integrations
Core	Started integration into Differ
Make sure the costs are as low as possible	Switched server provider
Improve loading times	Switched hosting provider, improved compression levels, reduced number of https requests, minified code
User manual	
Improve the user manual	Updated the user manual

Table 2. Identified Hubert needs and Actions taken after Pilot 1

Changes in Hubert after Pilot 2

From both the needs analysis and developers' own research with users, an increased need for additional intent training has been found so it was added to the improvement plan. Company was also aware that many students seem to appreciate the interface as a fun way of giving course/educator feedback, but also identified a need to make the experience more intuitive and more fun. Furthermore, developers revisited requests that concerned integrations with various other systems such as Differ, learning management software (LMS) and calendars. For the product in itself, most of the bugs that were encountered during Pilot 1 had been fixed and did not generate any new feedback points. Additional bugs were found in both the data export and in the results view. In terms of positive aspects in the user experience, many points that were identified in Pilot 1 were revisited in Pilot 2 where a large degree of users, both in terms of educators and students had previous experience with chatbots from varying settings.

- New requests were identified: more templates supporting more use-cases, additional input types for questions in the chat interface, improve the data export view (add timestamps and better structure), additional bug fixing.

Changes in Hubert after Pilot 3

The new learnings from Pilot 3 were the following:

Lots of different ideas about how the results page could be improved. The educators that we spoke to all had varying needs and no far-reaching uniformity could be found.

As a lot of feedback in Pilot 3 was related to the results page, developers put a lot of resources into perfecting it (see Figure 3). As the needs variated wildly, they rebuilt large parts of the system to support different needs.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

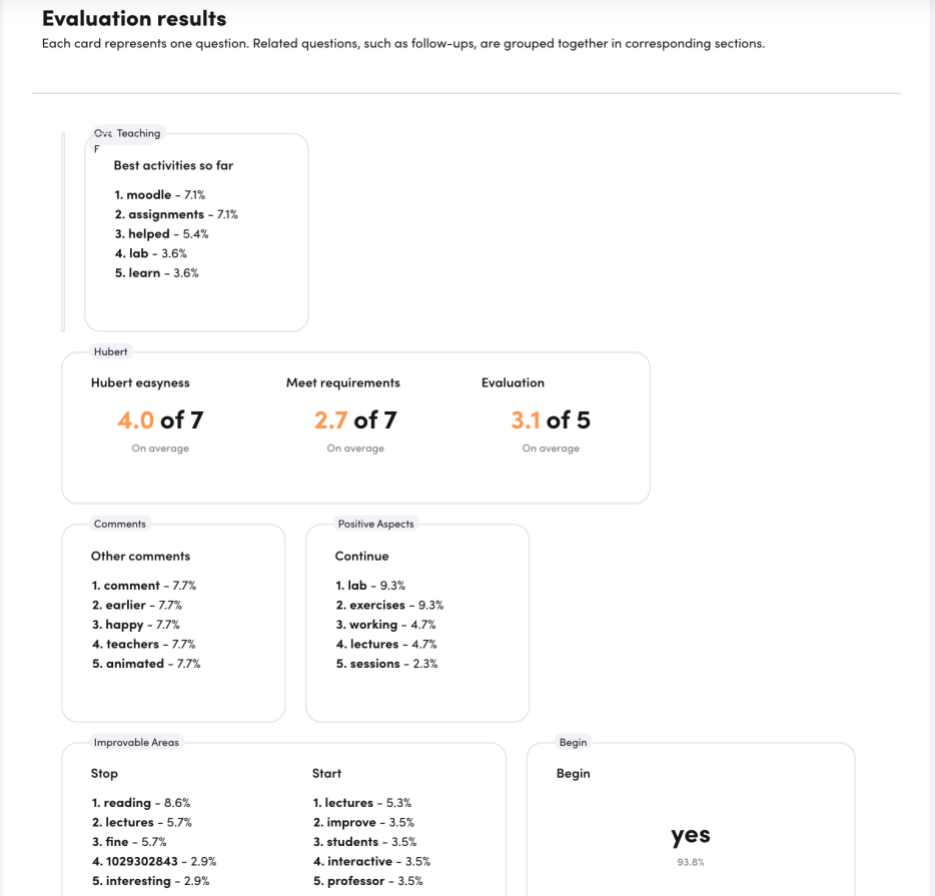


Figure 3. Updated Result view

It was also feasible now to support the use of Hubert inside Differ, but finishing the integration swallowed a large part of available resources as much of the existing back-end had to be rebuilt.

Company has also made a big effort to remove all bugs in the system. However, despite the integration with Differ, new requests for other integrations were received, which proved to be a major pain point for educators and seemingly restricted use due to the stand-alone system. Another interesting feature suggestion was to build a GUI for setting up questions templates individually. A great suggestion that surprisingly had not been brought up during previous pilots. The new results page seemed to please some of the audience, but many requests for added features were gained also during the fourth pilot.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

2. Pilot 1 courses

2.1. Differ - onboarding educators and students

Differ (https://www.differ.chat/) is a chatbot based application which is available as a mobile or desktop app. Users, students and educators also can use Differ through a web browser. Students sign up themselves for using Differ, educators just have to share a join-link with their students in an email or by linking it to LMS or Facebook group. Later, students receive SMS invitations to for example the course community, but their private mobile number is not visible to other users. It's completely GDPR compliant. Main advantage of the Differ is that it is a safe place for students to connect and have discussions with peers, it facilitates informal peer group conversations and easens student to educator communications. In addition to informal student communities, educators can use the Differ solution for academic purposes in course communities:

- Give students a personal welcome and open a line of communication.
- Empower the Student Voice to get continuous feedback from your students.
- Help students form study groups with peers in your class.
- Create conversations around Frequently Asked Questions.
- Connect and break the ice between students and student mentors/teaching assistants/student representatives or teaching staff.

Onboarding was the task of every subject educator, and invitation messages for students were sent through the learning management system (Moodle or Blackboard), email or in lecture. In total, in Pilot 1, Differ was used in 13 courses. Examples of inviting the students to join Differ in Pilot 1 courses are shown on Figure 4 and Figure 5.

Invitation message - COMP5840M Data Mining and Text Analytics:

Hello! 👋 Welcome to Leeds University. Please follow the link to join your peers on Differ.

COMP5840M Data Mining and Text Analytics

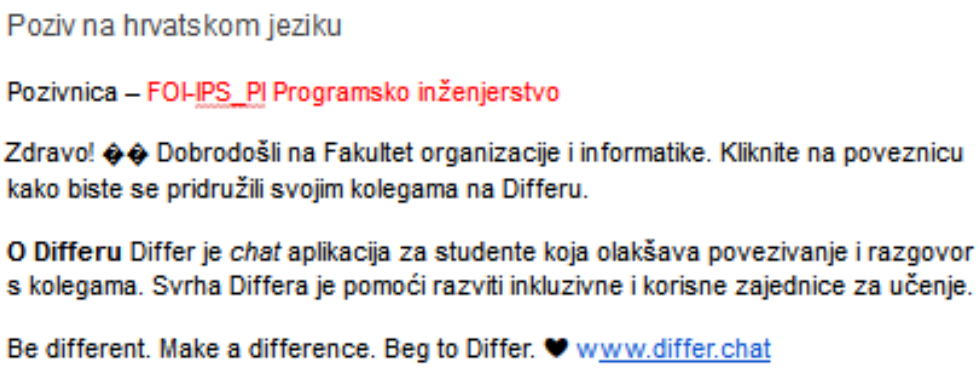
About Differ *Differ is a chat app for university students, making it easier for you to connect and chat with peers. The purpose is to help create inclusive and helpful learning communities.*

Be different. Make a difference. Beg to Differ. ❤ www.differ.chat

Figure 4. Invitation message in English language

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN


Poziv na hrvatskom jeziku

Pozivnica – FOI-IPS_PI Programsko inženjerstvo

Zdravo! ✿✿ Dobrodošli na Fakultet organizacije i informatike. Kliknite na poveznicu kako biste se pridružili svojim kolegama na Differu.

O Differu Differ je *chat* aplikacija za studente koja olakšava povezivanje i razgovor s kolegama. Svrha Differa je pomoći razviti inkluzivne i korisne zajednice za učenje.

Be different. Make a difference. Beg to Differ. ♥ www.differ.chat

Figure 5. Invitation message in Croatian language

How to use Differ?

For successful onboarding both educators and students, some of the partner institutions have made detailed guidelines (manual) to provide support for all potential Differ users. Others have relied on one-to-one online meetings with educators in Pilot 1 and written instructions via email. In the following paragraphs, we present approaches used by two partners.

Instructions for Differ use (Cyprus universities)

In pilot 1, CYENS invited educators across different local universities to participate in the pilot, for testing Differ. An invitation to join the pilot was shared with CYENS members' professional networks via emails. Specifically, the educators were provided with short and general descriptions for the potential use of Differ, as given below.

Differ: Developed to facilitate/support communication between students/learners in a mentor-mentee format

Educators who were interested in participating in the pilot were required to complete a short online form (i.e., Google form, https://form.jotform.com/200351810386448) and provide their contact details. Then, CYENS researchers then proceeded with communicating with the educators via phone or personal emails, for providing more information about the pilot, the potential use of the chatbots in the class, and the data collection processes.

Educators who were willing to proceed with the pilot testing Differ in their courses, further received instructions. The Differ community for their class was created by ETH members; the link to join the community was then shared with the educator, who subsequently invited the students in the same way. One-to-one tutorials and teleconferencing meetings took place, for providing support during the process. Also, written instructions were also provided on how to use Differ via emails.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Whole instructions English - Greek version are available in the document on this link and were used from Pilot 2 onwards.

Instructions in Croatian for students

In the first pilot round, educators from the Faculty of Organization and informatics created detailed instructions for students. The whole document ***WP5_UniZg_Differ user tutorial in Croatian.pdf*** is available on the link. **Instructions for Croatian students** involved steps regarding downloading the Differ, joining the course community on mobile, desktop or web browser version of Differ. Following section explained how to start private and group conversations in Differ, creating topics, annotating participants or highlighting the content in Differ. Figure 4. and 5. show part of these instructions.

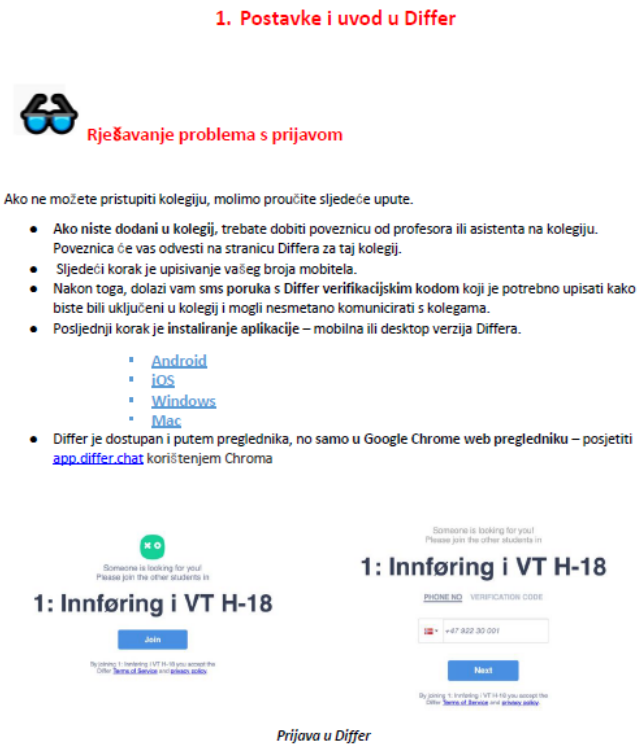


Figure 6. Differ sign-up

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

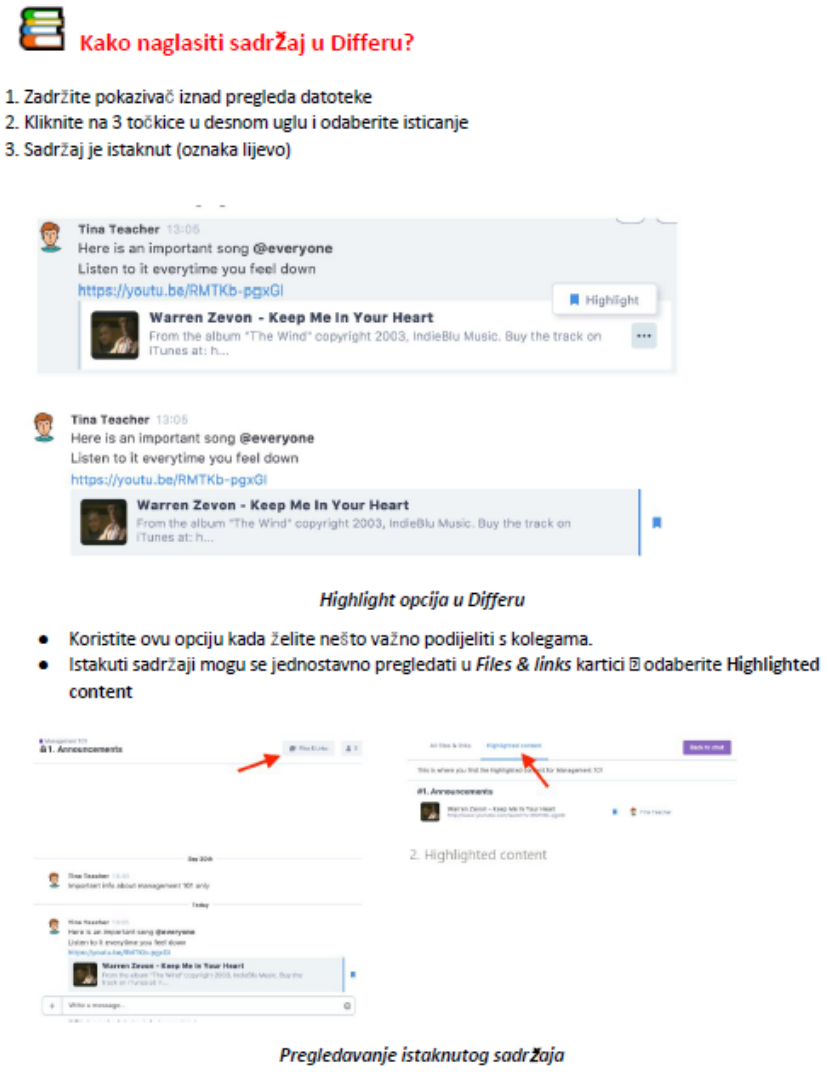


Figure 7. Highlight option in Differ for relevant topics

2.2. Hubert - onboarding the educators

Application Hubert (https://hubert.ai/) was used to perform evaluations of the Pilot 1 courses. This was done by using either the Education template *Start, stop, continue* or using the modified aforementioned template adapted for the EDUBOTS project's purpose. Template is a short survey that contains mostly open-ended questions to which students give their answers in conversational form, rather than choose from the offered answers. Questions are based on the modified Education template offered in Hubert (https://hubert.ai/).

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Team from FOI prepared for EDUBOTS partners instructions on how to implement the survey in Hubert and which questions are implemented in the EDUBOTS template (see Figure 8 and Figure 9).

Instructions

Each partner should prepare an environment for implementation of the surveys. We suggest the following:

1. Create and access your Hubert account, choose *Create new* from the menu, choose the custom EDUBOTS template and enter course details and invitation message before sending out the generated link. See Example 1 for the survey questions that are implemented in the EDUBOTS template in Hubert.

Figure 8. Instructions how to create a survey in Hubert

Example 1 – Questions in Hubert

Questions (implemented in the EDUBOTS template):

Hi, how old are you? Please type your age as a number.

What is your gender? Please choose Male or Female.

Chatbot offers answers: Male Female

What is working well on the course (course_name) and should continue in the same way?

What could teachers start doing, that would improve it?

What could teachers stop doing, that would improve (course_name)?

What is your overall experience of (course_name)? Please write a sentence or two.

Do you want to add something more?....

Figure 9. Questions in the template EDUBOTS

Each partner could customize the instructions and provide additional guidelines for their educators.

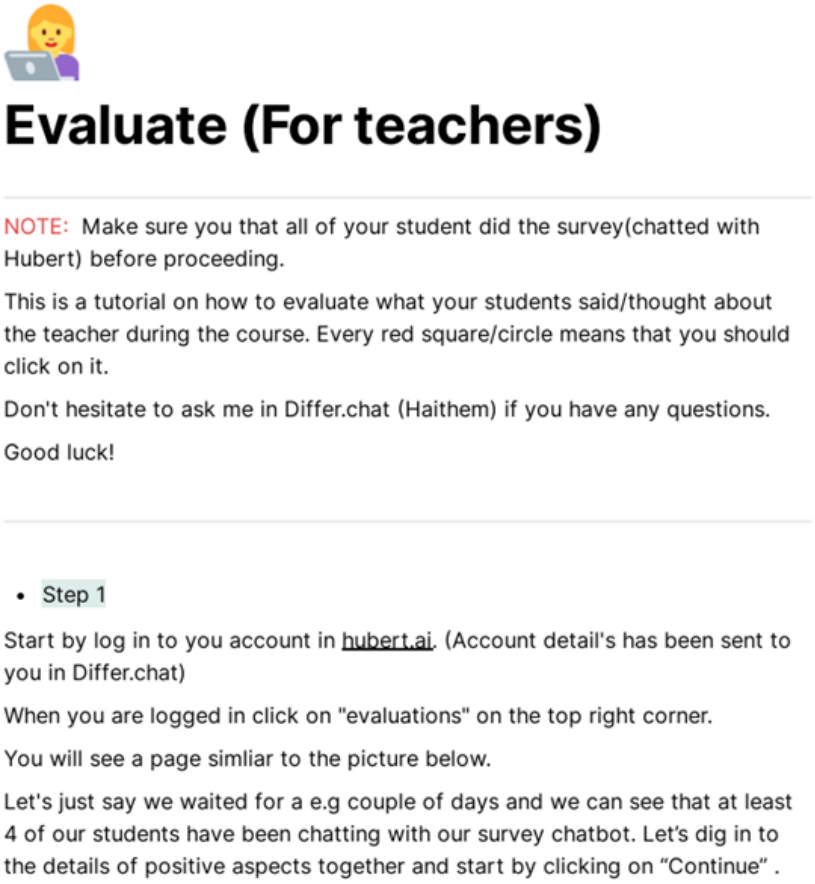
In pilot 1, CYENS invited educators across different local universities to participate in the pilot, for testing Hubert. An invitation to join the pilot was shared with CYENS members' professional networks via emails. Specifically, the educators were provided with short and general descriptions for the potential use of Hubert, as given below.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Hubert: Developed to support course evaluation in a formative manner (the Chatbot gathers input from the students in the form of a mini-chat interview with them at different time-points in the semester; a thematic map of the input is shared with the instructor ONLY).

Educators who were interested in participating in the pilot were required to complete a short online form (i.e., Google form, https://form.jotform.com/200351810386448) and provide their contact details. Then, CYENS researchers then proceeded with communicating with the educators via phone or personal emails, for providing more information about the pilot, the potential use of the chatbots in the class, and the data collection processes. Educators who were willing to proceed with the pilot testing of Hubert in their courses, further received guidelines and instructions on how to set up the chatbots and how to engage their students. A manual created by HUBERT was shared with educators in pilot 1 (see Figures below). In addition, one-to-one tutorials and teleconferencing meetings took place, for providing further support. Written instructions were also provided when needed on how to use Hubert, via emails.



Evaluate (For teachers)

NOTE: Make sure you that all of your student did the survey(chatted with Hubert) before proceeding.

This is a tutorial on how to evaluate what your students said/thought about the teacher during the course. Every red square/circle means that you should click on it.

Don't hesitate to ask me in Differ.chat (Haithem) if you have any questions.

Good luck!

- Step 1

Start by log in to you account in hubert.ai. (Account detail's has been sent to you in Differ.chat)

When you are logged in click on "evaluations" on the top right corner.

You will see a page simliar to the picture below.

Let's just say we waited for a e.g couple of days and we can see that at least 4 of our students have been chatting with our survey chatbot. Let's dig in to the details of positive aspects together and start by clicking on "Continue" .

Figure 10. Example of Hubert manual for teachers

The complete manual can be found at this link.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

2.3. Best practices in using chatbots – selected course case studies

Based on the data that were collected from the project partners, about the practice of using chatbot applications Hubert and Differ (see document Case studies), three case studies from Pilot 1 were chosen to represent the best practices in using the educational chatbots in the university courses. Those three courses applied different strategies in using chatbots, but that resulted in a great amount of interactions in the applications, particularly in the application Differ which was used to boost students' connectivity and interactions.

2.3.1. First case study – Software Engineering course (FOI UniZG)

Course description

Software engineering is a course held in the third year of undergraduate study programme at the University of Zagreb (Croatia), Faculty of Organization and Informatics (FOI), Department of Information Systems Development. The course is obligatory for students of the Information Systems study programme, and elective for students of the Business systems study programme. Students who enrol the course are mostly full-time students without or only a scarce knowledge of Software Engineering. The course is held in a summer semester and during the academic year 2019/2020 there were 230 students enrolled in the course. The course is provided by two professors and two assistants delivering lectures and computer exercises. Mean grade of the course (GTA) in the aforementioned academic year was 3.42 (grades in Croatian educational system range from 1-5). Out of 230 students, 50 students or 21.7% failed to complete the course successfully, meaning they need to enroll in the course once again.

Students onboarding into Differ

Students were invited during the first lecture and later an invitation and link to the Differ community FOI-IPS_PI were sent through LMS Moodle. The overview and benefits of using Differ were presented to students during the first lectures. The participation in Differ groups was not obligatory. One student, who already passed the course, had the role in the course as the student demonstrator (that is the usual role in that course), and he also took the role of a student mentor in Differ. However, the role of the student mentor was minor in this Differ setup. The focus was on student to student and student to educator communication.

The model of communication was developed in accordance with the course teaching model. Thus, we included:

- Educator to student communication through the formal topic for Announcements. This topic was closed for chat.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

- Student-student or mentor-student communication through the specific topics that were created for each topic the students have been preparing for their partial exams. The chat was very vivid in these groups.
- Student to educator communication through specific topics prepared specifically for the purpose of asking questions or having open chat related to the course.
- Student to student communication in private teams/groups/topics that were created by students.

Altogether, 183 students and 3 educators were onboarded into Differ. Out of 183 students, 118 of them (64.4%) have sent at least one message. Overall, there were 317 messages in all topics in the course community FOI-IPS_PI. There were 2338 messages exchanged between the members of the FOI-IPS_PI community, which is an average of 15 messages per user. Those messages included private direct messages (1931), private group messages (90) and topic messages (317). Altogether, there were 50.613 total sessions initiated in Differ, which makes this Differ community the most active one in the first Pilot.

Assessment of Differ

The general opinion is that a formal yet direct and easy channel of communication that was introduced to the course was a positive side of this "experiment". Due to the specifics of the course, the students were motivated and engaged in discussions in different groups. However, due to high expectations of such tools, the feedback from students and educators on user experience for using the tool was not very good.

25 students answered the questionnaire about usage of Differ. Majority of them (60%) used Differ several times a week. They mostly think that using Differ didn't help them in achieving a better grade (44% think that it didn't help, 24% are neutral) or in connecting with their peers (45% think it didn't help, 28% are neutral) but they think that it helped them in connecting with the educator (84%) and in active participation and successful completion of the course (64%).

Usage of Hubert

Hubert was used once at the end of the semester when all teaching activities were finished and students got points for the majority of their activities in the course, but they still did not have the final grade. Hubert was used to evaluate the course, not the educator's performance. 24 students participated in the course evaluation.

2.3.2. Second case study – Text and Image Processing Course (FOI UniZG)

Course description

Text and Image Processing is a course held in the 2nd semester of the first year of undergraduate study programme at the University of Zagreb (Croatia), Faculty of Organization and Informatics

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

(FOI). The course is obligatory for students of the Information Systems study programme. Students who enrol the course are mostly full-time students with different knowledge of text and image processing, based on their previous secondary school level programme. The course is held in a summer semester and during the academic year 2019/2020 there were 209 students enrolled in the course. The course is held by two professors and three assistants delivering lectures, auditory exercises, and computer exercises. Mean grade of the course (GTA) in the aforementioned academic year was 3.6 (grades in Croatian educational system range from 1-5). Out of 209 students, 48 students or 22,97% failed to complete the course successfully, meaning they need to enrol in the course once again.

Students onboarding into Differ

Out of 5 educators working on the course two were actively engaged in Differ. They onboarded 138 (66%) out of 209 students. Since the course has over a dozen groups for computer exercises Differ was primarily used to support complex work during the exercises which were held online because of the COVID-19 pandemic. The onboarding was done via invitation which was sent through the learning management system to student's emails. Joining Differ was a part of one of the non-obligatory computer exercises tasks. In the task students were asked to create the profile and shortly introduce themselves in the Differ group for their computer exercises.

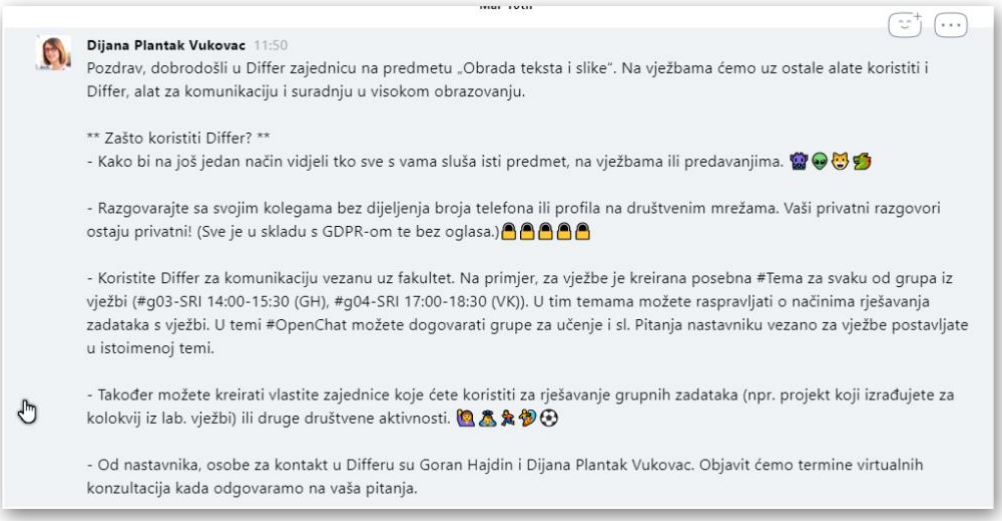


Figure 11. Welcome message in Differ for the course Text and Image Processing

Students were asked to suggest a colleague to be a mentor student in Differ or to apply by themself, and then "vote" with like. Educators motivated students to take the role of the student mentor in their laboratory groups by offering extra course points for engagement in Differ. Approximately one third of the laboratory groups had student mentors. Despite all efforts several groups were left without student mentors. In those groups educators assumed the role.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

As a part of the laboratory task, a brief overview of Differ was given emphasising main benefits. To ruther support use of Differ a manual in Croatian was created (used in each FOI course) based on the Differ help documentation.

The initial plan was to use Differ during computer exercises in physical classrooms, using a mobile application or a web browser, so that students could share their tips and solutions. Initial plan was realised only for the several groups in the first usage of Differ and students were actively participating. Due to the COVID-19 pandemic all teaching activities shifted fully online. Educators then initiated communication in Differ for each of their groups each week by explaining exercises tasks and gave support to all student groups. Communication with student mentors was mostly in a private channel in Differ. Some of the student mentors provided tips and solutions to the problems or initiated communication in their respective group. Myriad of students asked questions regarding particular exercise.

Assessment of Differ

Despite a lot of engagement of the educators to motivate the communication, there were no responses in some laboratory groups, just one-way communication. Since communication in Differ was less formal than in official channels (LMS, email), educators perceived that communication with the students was more frequent than in previous years.

19 students participated in the online questionnaire regarding usage of Differ. Majority of them used Differ once a week. 47,4% of participants agreed that using Differ helped them to finish the course successfully and 68,4% answered that using Differ helped them get a better grade in the course. Differ helped them to connect with other students (57,9%) or with the educators (46,3%). Differ helped them to actively participate in the course and to successfully complete the course activities (52,6%), but not so much for collaboration with other students (36,8% agreed and disagreed on using Differ for collaboration).

Out of 138 students who onboarded into Differ, 119 participated at least once in the groups. Average message per user was 4 and average sessions per active user was 76. This shows that students were using Differ, but mostly passively, to receive information from the educator or colleagues.

Usage of Hubert

Hubert was used once at the end of the semester when all teaching activities were finished and students got points for the majority of their activities in the course, but they still did not have the final grade. Hubert was used to evaluate the course, not the educator's performance. 29 students participated in the evaluation.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

2.3.3. Third case study – Design for all (CYENS/UniCyprus)

Course description

The course Design for All is an obligatory course in the scope of the undergraduate study program Multimedia and Graphic Arts at the Cyprus University of Technology (CUT). It is held at the third year of study, in the 6th semester. The course started in mid January 2020, took place physically until the beginning of March 2020 when all classes switched to online mode due to Covid-19. The course was held by one lecturer, and 39 students attended. All of the students have completed the course successfully. Mean grade of the course (GTA) was 7.5 (grades at the CTU range from 0 to 10).

Students onboarding into Differ

Since the course had a relatively small number of students, only one lecturer was involved with the course and was piloting Differ in the scope of the course. Students were onboarded by email. Number of onboarded students was 30. A teaching assistant, a senior student, and an EDUBOTS researcher were onboarded as mentors.

After onboarding, students were greeted with the message presented below in italic.

Hi! Welcome to this Differ community for course MAG320 Design for All at CUT.

My name is Olia and I'm the community moderator, on behalf of [NAME OF THE INSTRUCTOR BLINDED]. Don't hesitate to reach out to me if you have any questions. (You can send me a private message).

Why should I use Differ?

- See who's in your course. 👾👽🐱🐲

- Chat with your peers without having to share phone numbers and social media accounts.

Your private conversations stay private! (GDPR-compliant & Ad-free) 🤳🤳🤳🤳🤳

- Use Differ for course communication with your peers. For example, you can ask questions in the #OpenChat and help your peers by answering. Also, you can create your own communities to use for group assignments or other social activities. 🙋🏕️🏃🌮⚽

[NAME OF THE INSTRUCTOR BLINDED] will have virtual office hours on Differ, where she will be answering any questions. But he won't hang out on Differ 24/7, so in case you want a quick answer - ask the question in the #Open Chat and get help from your peers. (And of course, help whenever you can)

About Differ

Differ is a chat app for university students, making it easier for you to connect and chat with peers. The purpose is to help create inclusive and helpful learning communities.

Be different. Make a difference. Beg to Differ. 💔

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

The mentors tried to initiate discussions with the students on a personal level. However, they unofficially admitted that the communication and response on the behalf of the students was limited. Once the community was set up and students and mentors were onboarded, there were some interactions at the beginning, at least in the open chat space.

Assessment of Differ

The only available data comes from the post-pilot survey in which the students were asked to indicate frequency of Differ usage. The survey was administered by all of the participants, thus there is no data per course. 32 students responded to the survey, out of which 14 students did not use Differ at all and thus they did not complete the survey. From the rest of the students (n=18), one student used Differ twice, one student used it 11-20 times, and 16 students used it 1-5 times during the semester.

Total of 16 Differ users sent messages, and the number of total messages sent is 123. Out of those 123 messages, 30 were private direct messages, 69 were private group messages, and 24 messages were messages in all topics.

Usability of Differ was measured with the System Usability Scale (SUS). The mean SUS score is 67.78±15.19, with a minimum value 25.0 and a maximum value 85.0. A SUS score above a 68 would be considered above average and anything below 68 is below average. That means that the perceived usability of Differ is average.

Usage of Hubert

The students were prompted twice in the semester to use Hubert for course evaluation purposes.

2.4. Interactions in Differ and Hubert

2.4.1. Analysis of system data in Differ

Transcripts of interactions which took place on the Differ platform during the project's pilot phases were not available, as this would be at odds with users' expectations of privacy.

The primary sources of Information about the ways in which the Differ platform was used were the pilot participant surveys, interviews and discussions, as well as anecdotally by observation of the Differ communities in which consortium members participated.

High-level aggregated data on the volumes of user activity (e.g. logins, messages sent) by Differ community were collected by Edtech Foundry (ETF) and made available to consortium members. The summary data includes frequency counts for a variety of events recorded by the application such as sessions started, messages sent and so on. These were aggregated by community, pilot phase and day. In each case, a daily total count and a count of unique users completing the action were supplied. A count of student users who were ever active in each community and pilot phase

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

was also obtained, in which was included any (anonymous) user ID for which any activity other than “session started” was observed.

The Differ communities created during the project were open to students from a range of different populations. They can be categorised into four community types:

- Course – covering individual modules or units of study. All students in the community were likely to be consuming the same materials, undertaking the same assessment etc.
- Programme – students across a larger programme of study, e.g., a particular MSc course. Students in these communities are likely to have had substantial overlap in modules undertaken, academic interests and so on.
- Department – covering multiple programmes within the same school or department, likely across multiple years of study (e.g. all undergraduate Computing students).
- Multi-Disciplinary – more disparate groups of students, with something in common other than the subject they are studying. For instance, communities open to all postgraduates, students affiliated with CYENS’ MiHub programme for vulnerable migrants[1], or with initiatives aimed at widening participation in HE[2].

The number of communities created and volume of users ever active in Pilot 1 are shown below.

Community Type	Pilot 1 Communities	Users Ever Active by Differ Community		
		Minimum	Maximum	Mean
Course	9	2	99	35.4
Programme	0	-	-	-
Department	1	9	9	9.0
Multi-Disciplinary	1	5	5	5.0
Total	**11**	**2**	**99**	**30.3**

Table 3. Number of communities and active users in pilot 1

2.4.2. Analysis of data from the course evaluations in Hubert

Transcripts of student interactions with the Hubert chatbot were made available during the project. These transcripts were ingested and transformed to enable downstream analysis of their contents while taking steps to protect the privacy of educators and ensure that the data were consistently cleaned and made user-friendly.

These ingestion and transformation steps are described in detail in the project Data Report, but included the exclusion of irrelevant data (e.g. test conversations and those where students

[1] https://mihub.eu/en/
[2] E.g. https://students.leeds.ac.uk/info/1000014/plus_programme/1066/about_the_plus_programme

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

indicated that they had not taken the course under evaluation) and translation of non-English responses into English.

See section 5.2.2. Analysis of data from the course evaluations in Hubert below for analysis of Hubert activity across all pilot phases.

2.5. Results of the surveys about Differ

2.5.1. Differ and learning process

In the line with the goals of EDUBOTS project, we employed the survey to explore the role of the application Differ from two standing points: (I) to increase student engagement in course and reducing dropout rates and (II) to help students in bonding with peers and educators to develop social connections and sense of inclusion to the faculty.

The online survey was created in Google Forms and participation was voluntary and anonymous. The approval of the Ethics Committee was obtained before carrying out the pilot study. First part of the online survey consisted of demographic questions of participants: age, gender and course in which they have used Differ. Second part was devoted to seven items related to frequency of using chatbot application, learning process and course participation (e.g. “Using Differ helped me to actively participate in the course and successfully complete all activities.”) and communication with peers and educators (e.g. “Using Differ helped me to connect with the course educators.”). A five-point Likert scale (from “1 – I do not agree at all.” to “5 – I fully agree.”) was used to assess the respondents’ level of agreement with the statements. One open-ended question was included so that students could describe their experiences and observations in more detail: “If you want to list something else related to how Differ helped you in the learning process, please do so here.”

Respondents of the Pilot 1 study were 76 students of Faculty of Organization and Informatics (n=76), who evaluated the role of chatbot-based communication platform Differ for academic purposes in the summer semester of 2019/2020. academic year. Data was gathered on five courses at undergraduate level of study. Regarding demographic characteristics of the sample, 40,79% male and 59,21% female evaluated Differ.

A first version of Differ was employed, in which focus is put on the facilitation of communication between educators and students in order to enhance teaching and learning processes. Students were informed about Differ by their educators who sent a brief description of Differ, an invitation link to the respective community, and a manual in Croatian language based on the Differ help documentation. Differ invitations and guidelines were sent through the LMS Moodle to the students’ email addresses.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Interconnection of successful course completion, engagement and Differ use

Regarding frequency of use, most of the students used Differ several times to once a week. Only two students used Differ every day, and only 19,76% of respondents were prone to monthly use of the application within the course.

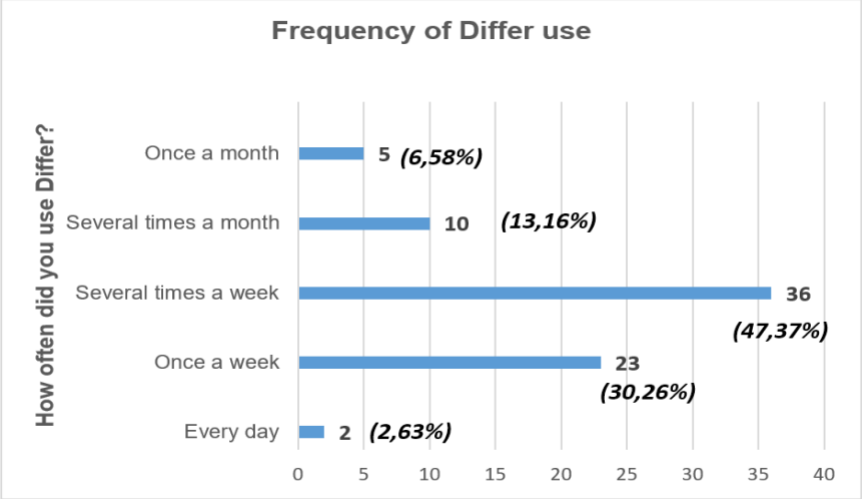


Figure 12. Frequency of Differ use in a pilot study, n=76

From analysis of students' responses to the three items related to **course performance and grade achievement,** following results were obtained:

- Students evaluated Differ very positively as a supporting tool for active course participation and successfully completing the course.
- Half of the students (50%) reported that Differ helped them in achieving better grades and the other half was indifferent or disagreed with this statement.
- 55,26% of the students considered Differ as a tool that has helped them for successful course completion and even 61,84% of them recognized Differ's role for supporting active participation and finishing course activities.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

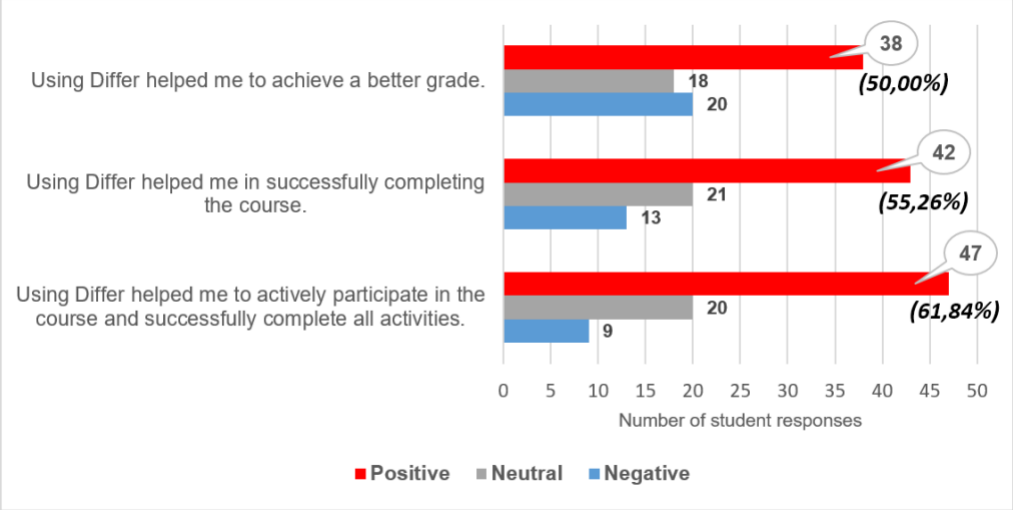


Figure 13. Student responses related to course performance, n=76

To examine the relationship between the frequency of use of the Differ application and student academic success, Spearman correlation coefficient was calculated and results are shown in Table 4.

Learning engagement items	**Spearman's correlation coefficient**
Using Differ helped me to achieve a better grade.	.141
Using Differ helped me to actively participate in the course and successfully complete all activities.	**.327****
Using Differ helped me in successfully completing the course.	**.330****

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4. Results of correlation analysis of frequency of using Differ and learning engagement items

We can conclude that in the case of this pilot study, more frequent use of Differ was correlated with slightly increased participation and completion of course activities and more successful course completion among students.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

The role of the Differ application in connecting students with colleagues and educators in the course

From the interpretation of student's answers from the survey (Fig. 5), following results were obtained:

- Differ did not have a great positive impact on connecting students. Only 34 students, or 44,74% of them, think that Differ helped them to connect with other students. Others have a rather negative or neutral attitude.
- Chatbot application was also not recognized as particularly useful for collaboration on the course-related activities (63,16%, or 48 of them, disagreed with the statement or were indifferent)
- On the other hand, Differ made a **significant impact on student-educator connections. 84,21% of students agreed that using Differ helped them to connect with the course educator(s).**

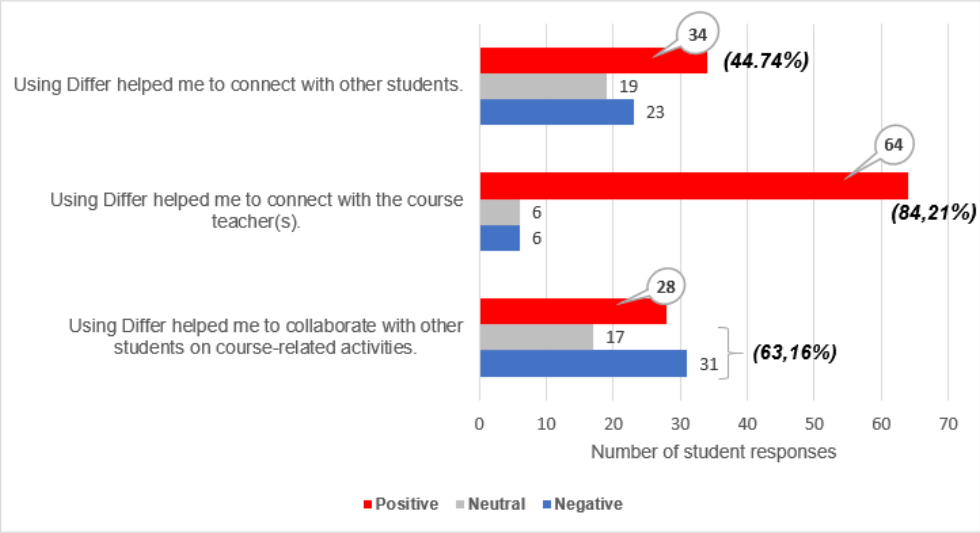


Figure 14. Differ's role in peer-to-peer and educator's connections

From Pilot 1 study experience, Differ can be successfully implemented in various course contexts and deal with a large group of studentsSince the students in pilot study evaluated strongly positive chatbot applications for improving communication with course educators, Differ can be considered as a safe environment where students can control and express their thoughts without being judged. Promoting Differ as a secure place is one of the it's pillars in strengthening social inclusion in the academic context. Students from our study revealed generally positive or neutral attitudes towards chatbot. However, even though the Chabot application gained a positive impression because of establishing a more approachable and informal communication channel with course educators, **results show lack of peer-to-peer communication and collaboration habits on the Differ platform** for course activities or achieving better course grades.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Extended report about students' engagement in Differ is provided in the research paper by Čižmešija, Ana; Horvat, Ana; Plantak Vukovac, Dijana (2021). Improving student engagement and course completion using chatbot application. In: Gómez Chova, L., López Martínez, A., Candel Torres, I. (eds.) Proceedings of INTED2021 Conference. IATED, Spain, pp. 8346-8354, https://doi.org/10.21125/inted.2021

2.5.2. Usability and User Experience of Differ

Previous research on the usability and user experience (UX) of chat platforms and chatbots used in education is scarce, so it was worthwhile to explore Differ's quality in use.

In order to gain a comprehensive user view of Differ aligned with the EDUBOTS project goals, the research team has decided to evaluate its usability and user experience by collecting opinions from the users, primarily students, about their satisfaction with the application. In the past few years, User Experience Questionnaire (UEQ) has been found to be frequently used for UX evaluation, with the complementary method System Usability Scale (SUS) as the most common one for usability assessment (Díaz-Oreiro et al., 2019). System Usability Scale is a highly robust and versatile usability ten-items measurement tool which has been used among researchers and industry experts for almost 30 years (Bangor et al., 2008; Hassenzahl & Tractinsky, 2006). User Experience Questionnaire (UEQ) is a tool that measures both usability and hedonic quality aspects by applying 26 pairs of contrasting attributes (Schrepp et al., 2017; *User Experience Questionnaire (UEQ)*, n.d.).

For this research, ten items of the original SUS scale with answers on a five-point Lickert scale, and 26 pairs of attributes from the UEQ were implemented into the Google Forms online survey and a link to the survey and research invitation was sent to the students via the course learning management system.

Total of N = 76 students from the Faculty of Organization and Informatics, who have used Differ in five courses in the academic year 2019/2020, completed the online survey. Regarding demographic characteristics of the sample, 40,79% male and 59,21% female evaluated Differ.

Usability of Differ

Students' opinions about Differ's ease of use were analysed and revealed that the mean SUS score was **72.4** (StDev = 0.8) and the median was 76.3. The highest score was 100.0 and the lowest was 27.5. A boxplot of Differ SUS scores is shown in Fig. 1. This positions the evaluated version of Differ as **slightly above average application**, given that all SUS scores above 68 would be considered above average. In relation to other descriptive ratings [34], Differ is a "good" application (adjective rating) earning a school grade C + and having an acceptable interface.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

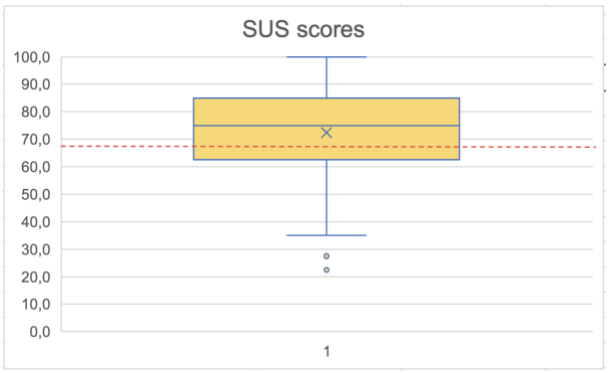


Figure 15. Differ SUS scores (Benchmark of 68.0 is marked in red dotted line)

User experience of Differ

Items of the User Experience Questionnaire (UEQ) are grouped into the 6 scales: Attractiveness (gives overall impression about the product), Perspicuity (measures learnability of using the product), Efficiency (measures efficiency of interac- tion with the product), Dependability (measures levels of control of the product), Stimulation (measures levels of motivation to use the product) and Novelty (measures levels of innovation or originality of the product) (Schrepp et al., 2017). Each scale ranges from -3 (horribly bad) and + 3 (extremely good) values, corresponding to the positive or negative attribute on a differential scale, e.g. on the Attractiveness scale a product is described with the pairs unpleasant-pleasant. Values above 0,8 represent a positive evaluation of the quality aspect, while values below 0,8 represent a negative evaluation.

Analysis showed that Differ is **positively evaluated** on the scales *Attractiveness*, *Perspicuity*, *Efficiency* and *Dependability*, but achieved lower scores on the scales of *Stimulation* (mean 0,438) and *Novelty* (mean 0,572), which is shown in the Fig. 2. Two later scales, Stimulation and Novelty, which are included in the hedonic quality of Differ, were negatively assessed with a mean of approximately 0,5.

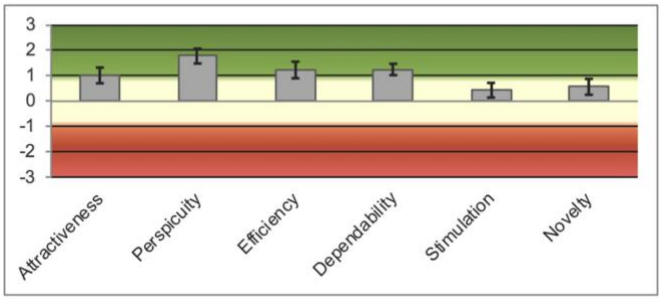


Figure 16. User experience of Differ measured by UEQ scales

Lower scores on the hedonic quality aspect are further underpinned by a benchmark against a data set from 452 studies concerning different products (business software, web pages, web shops, social networks). Benchmark revealed (see Fig. 3) that ***Stimulation*** **for using Differ is**

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

low and also the worst aspect of the application (it belongs to the range of the 25% worst results). Furthermore, the ***Novelty*** **or originality of Differ is below average** (50% of results of other studies are better, 25% of results are worse).

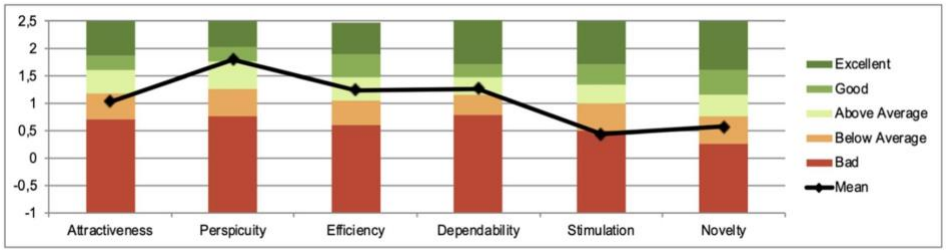


Figure 17. Positioning of Differ on a UEQ benchmark with 452 studies

We can conclude that Differ has been perceived as a quite ordinary chat application, which suggests that chatbot functionalities were not recognized or used in the application. Results of evaluation of Differ's user experience suggest that future redesign of the application should consider factors that stimulate students to use it frequently and give additional value to their academic life, which could be achieved by employing chatbot functionalities which are relevant for the educational environment.

Extended report about usability and user experience of Differ is provided in the research paper by Plantak Vukovac, D**.,** Horvat, A., Čižmešija, A. (2021) Usability and User Experience of a Chat Application with Integrated Educational Chatbot Functionalities, In: P. Zaphiris and A. Ioannou (Eds.): HCII 2021, LNCS 12785, pp. 1–14, 2021, https://doi.org/10.1007/978-3-030-77943-6_14

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

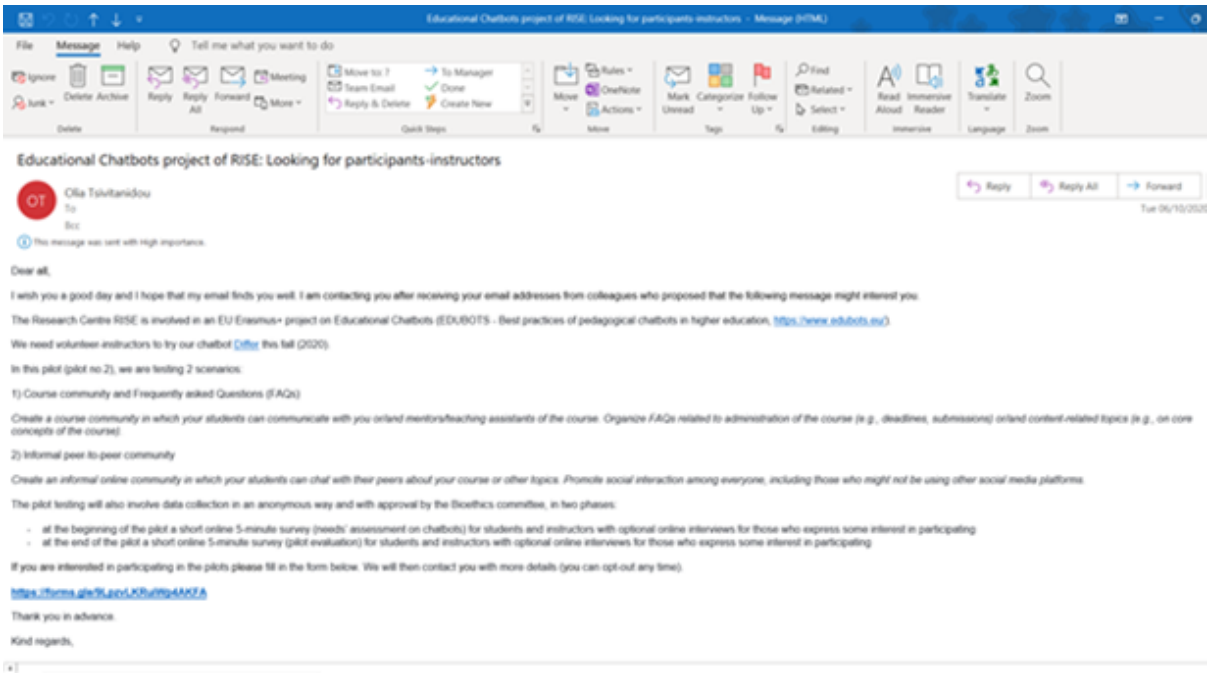
3. Pilot 2 courses

3.1. Best Practices in using Differ and Hubert – selected course case studies

3.1.1. Courses at the Cyprus Universities

Onboarding

Educators were recruited by an email invitation (see figure below) that was sent to our organisation, local universities, and our network of academics. Educators were briefly informed about the pilot, the two scenarios that we were mainly testing (FAQs, informal peer-to-peer community) and were asked to fill in a form, by clicking on a link (https://forms.gle/9LpzvLKRuiWp4AKFA) that was provided at the end of the email text, to express their interest to participate in the pilot.



Educational Chatbots project of RISE: Looking for participants-instructors

Olia Tsivitanidou

Dear all,

I wish you a good day and I hope that my email finds you well. I am contacting you after receiving your email addresses from colleagues who proposed that the following message might interest you.

The Research Centre RISE is involved in an EU Erasmus+ project on Educational Chatbots (EDUBOTS - Best practices of pedagogical chatbots in higher education, https://www.edubots.eu/).

We need volunteer instructors to try our chatbot Differ this fall (2020).

In this pilot (pilot no 2), we are testing 2 scenarios:

1) Course community and Frequently asked Questions (FAQs):

Create a course community in which your students can communicate with you or/and mentors/teaching assistants of the course. Organize FAQs related to administration of the course (e.g., deadlines, submissions) or/and content-related topics (e.g., on core concepts of the course).

2) Informal peer-to-peer community

Create an informal online community in which your students can chat with their peers about your course or other topics. Promote social interaction among everyone, including those who might not be using other social media platforms.

The pilot testing will also involve data collection in an anonymous way and with approval by the Bioethics committee, in two phases:

- at the beginning of the pilot a short online 5-minute survey (needs' assessment on chatbots) for students and instructors with optional online interviews for those who express some interest in participating
- at the end of the pilot a short online 5-minute survey (pilot evaluation) for students and instructors with optional online interviews for those who express some interest in participating

If you are interested in participating in the pilots please fill in the form below. We will then contact you with more details (you can opt-out any time).

https://forms.gle/9LpzvLKRuiWp4AKFA

Thank you in advance.

Kind regards,

Figure 18. Invitation e-mail

Once receiving the responses from educators, the next step for us was to contact the educators, and provide some more information about the pilot, either via email or phone, and if needed, arrange 1-to-1 online meetings with them, explaining the aims of the pilot, showcasing the two

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

chatbots (i.e., BO chatbot within Differ, Hubert) as part of educator training (T2.5) and then explain the scenarios that could be tested. The instructions provided to educators, for using Differ and Hubert, are provided below. Five educators expressed their interest in participating to pilot 2, while four of them proceeded with their participation; thus, we had a total of four courses participating in pilot 2, but a total of five use cases applied, as in one of the courses both Differ, and Hubert were tested (see case studies report for more information).

In general, Differ provides a secure communication environment between students and educators, mentors, teaching assistants, etc. The Differ tool can be used in two ways:

Option 1: Course community and Frequently asked Questions (FAQs)

Create a course community where your students can contact you and/or course mentors/ assistants. Organise frequently asked questions (FAQs) about course management (e.g. deadlines, submissions) and/or topics related to the course content (e.g., key course concepts).

Option 2: Informal peer-to-peer community

Create an informal online community where your students can talk to their classmates about your lesson or other topics. Promote social interaction among students, including those who may not use other social media platforms. In this case, you can act as the administrator of the community that you will create within Differ and where you will invite the students, have a passive role, leaving the students to use the tool on their own to communicate with each other.

Manual for Differ use (English-Greek version)

Instructions to educators for the use of Differ (in Greek; the instructions are also available in English) contain the following **Differ installation steps and utilization of basic features:**

Step 1: **Register in Differ**, via the page: https://join.edubots.eu/sign-up. Fill in your contact details and select your institution. Then press the button «Sign up».

Step 2: **Subscribe to the Edubots Educator community on Differ by clicking on the link:** dif.re/edubots. In this community you can find other educators using Differ. Download Differ (Desktop or mobile app).

Step 3: **Welcome to Differ.** You will first see BO, a chatbot with whom you will interact within Differ. BO will help you make the most of Differ's capabilities, such as creating your own community and creating your own FAQs (if take advantage of this opportunity). The image below explains the icons you will see in the Differ environment.

Step 4: **Choose «Educators, this is for you! ».**

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Step 5: Choose «Create a community for students».

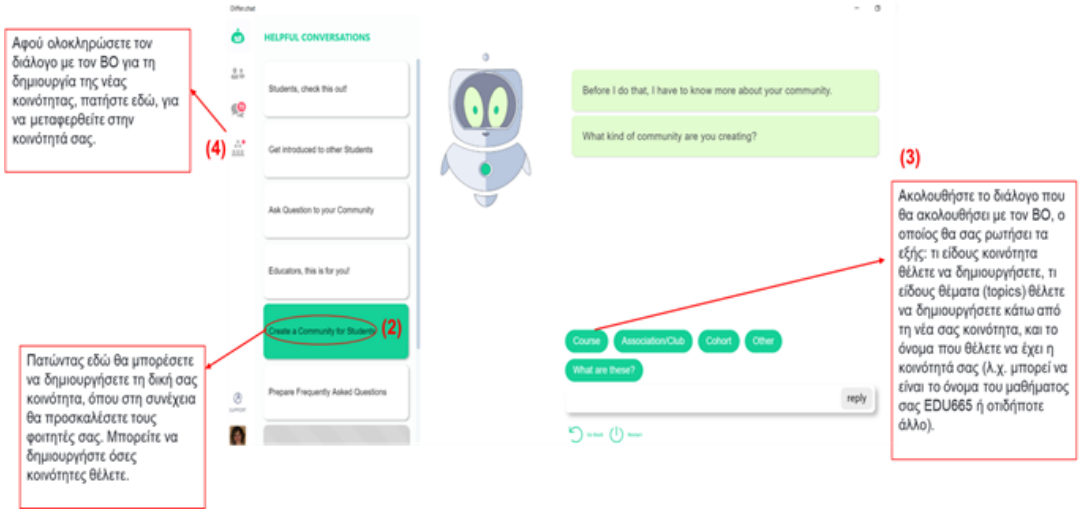


Figure 19. Differ instructions in Greek language - Create a community for students.

Step 6: Your community has been created. What can I do within the community?

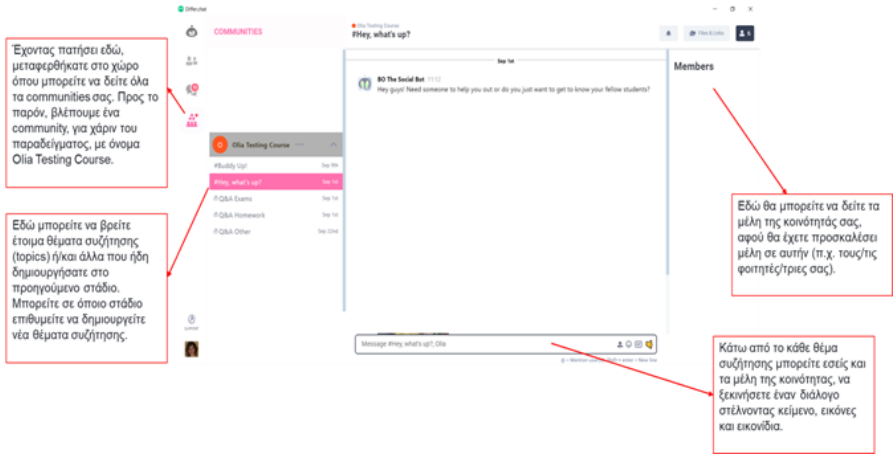


Figure 20. Differ instructions in Greek language - activities within community

Step 7: How do I invite students to the community I have created? How do I create a new topic within my community?

Step 8: How can I create Frequently Asked Questions (FAQs)?
Step 9: How can I find the FAQs that I created within my community?

Additional information is provided in The T2.5 ‘Educators’ training material’ WP2 report.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Instructions for creating Hubert evaluations

Link to the Hubert application is here. https://hubert.ai/. Below are presented steps for creating a Hubert account and how to start a Hubert evaluation.

Step 1: create your Hubert account

Step 2: login and go to evaluations

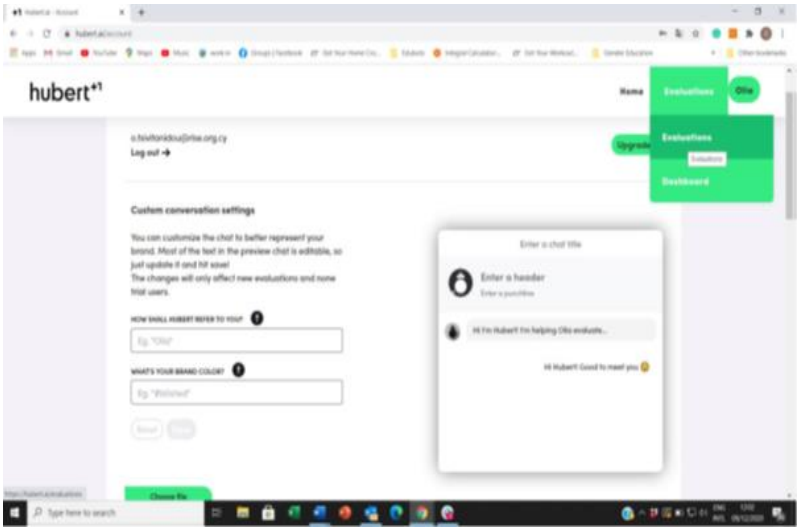


Figure 21. Hubert instructions

Step 3: choose “create new”

Step 4: choose domain “education”

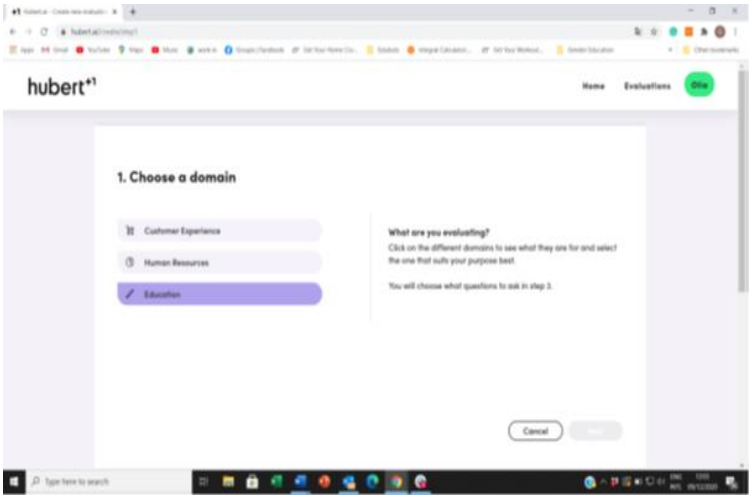


Figure 22. Hubert instructions

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Additional information is provided in The T2.5 ‘Educators’ training material’ WP2 report.

Best case (Cyprus University of Technology)

The following use case was one of the most successful cases of our pilots. The Mihub community was created in Differ in the first pilot of the project and was active and continuously used throughout the lifespan of the project, until the fourth pilot. The Mihub community was created to provide help and support to migrants from Non-European Countries in order to adjust to the local society, including familiarizing themselves with the local language. For Mihub members, the Differ community was valuable, especially during the pandemic period during which physical meetings were limited, for communicating with other members of the community, for discussing issues with the community moderator and for finding answers to commonly asked questions.

Course name: Language course for migrants (Mihub Information Center via the Cyprus University of Technology 2020-2021).
Number of students: 50 to 100
Course description: Help and support is provided to migrants from Non-European Countries in order to adjust to the local society, including familiarising themselves with the local language.
Description of students’ engagement with Differ and/or Hubert: The facilitator created an online community for Mihub, using BO, and invited Mihub members to join via email. Primary focus of Differ use on the course were the FAQs created by the facilitator, but also the facilitator-members and member-to-member communication through Differ, its use for announcements related to the topics of interest to migrants, and Q&A (ask questions to your community).
Affordances and challenges: Mihub members were interested in using Differ and the facilitator claimed that the online community helped its members to interact with people having the same needs and sharing similar concerns with them. Main affordance was that due to the COVID-19 pandemic and the mitigation measures (i.e., lockdown), the online community in Differ offered a safe environment for communication and exchange of information, when the circumstances did not allow face-to-face interactions. The facilitator claimed that he frequently used the FAQs feature, and it was useful for the Mihub members to find answers to those commonly asked questions. Also, the facilitator had the opportunity to make other postings and announcements in the open chat space, and many members of the community communicated with him via private messages. Differ was used on an everyday basis by the facilitator and the community members. The facilitator claimed that he would continue using Differ, even after the end of pilot 2.
Indicative quotes from the interview with the facilitator:

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Figure 23. Feedback from Differ users

Overall opinion of educators participating in Pilot 2

Some of the educators expressed their concerns about the added value of Differ/BO on their own teaching and/or students' learning, since Differ is mainly a chat tool, and many different chat tools or communication platforms already exist and are being used by the students. Especially in small scale classes with students who are on their 3rd year of studies onwards and being used to the use of other platforms and social media channels for communication and interaction, it was hard to find value for the use of Differ in class, either for establishing informal communities, or formal communities with the active involvement of the educator. Educators acknowledged though, that Differ could be a useful and safe environment to use for communication with students, especially for large scale classes, and for freshmen. The use of Differ and BO was a success in the Language course for migrants (Mihub Information Center via the Cyprus University of Technology 2020-2021). This community was comprised by migrants, people who share some common concerns, who seek for using media tools for interacting with other people of their community that they do not know already (and therefore cannot find easily in commonly used social media channels) and with a counsellor/ facilitator that can offer support in helping them learn the local language, familiarize themselves with the local system, and even helping them to find a new job. Thus, the fourth instructor found value in the use of Differ within this community, in which all members were actively involved. Especially during the COVID-19 pandemic and the lockdown, Differ offered an alternative channel of communication and interaction among MiHub members.

3.1.2. Courses at the University of Granada

Onboarding

Taking into account the low success of Pilot 1 at UGR, in Pilot 2 we changed the way to enrol or onboard people. First, a public call for recruiting the so-called 'mentors' was launched for the whole University, however, it was focused on the Informatics and Telecommunication School of Engineering (ETSIIT). 8 students were contracted (they were paid) as mentors, 4 belonging to

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

the Degree in Computer Sciences and other 4 studying the Degree in Telecommunications Technologies. Their tasks included introducing Differ application and chatbot to their peers, trying to create different communities at ETSIIT inside Differ. They used standard channels to communicate with other students (e-mail and Telegram lists, Student Delegation advertisement, etc). There were created some general communities, the most prolific one was focused on post-graduate students, but none of them was very active.

A webinar about the use of Differ was organised by the EDUBOTS project consortium (in September 2020). We advertised the webinar in the webpage of the International Projects, Networks and Agreements Office from UGR, as well as through the e-mail lists of the ETSIIT center. They were interested in the technology and wanted to learn about the possibilities of Differ for education. Several educators attended the webinar, however, they didn't find the tool useful to be used in their classes. In addition we sent some e-mails to the ETSIIT educators (in the School of Engineering mailing list), trying to enrol some of them to use Differ in their classes, but this was not also successful.

So, finally, only the educators being part of the project team applied Differ or other tools in their subjects (3 in total), and only the students of their courses participated in the pilot experience.

Best case

This use case was the most successful in Pilot 2 using the Differ application and chatbot BO. It was conducted by one of the EDUBOTS educators from UGR in one of his subjects. The description of the case is below.

Course name: Fundamentos de Redes (Fundamentals of Networking).
Number of students: 50 to 100
Course description: It is provided to students at their third year in bachelor's degree studies in Computer Sciences. The course presents the fundamentals of Computer Networks, including the conceptual OSI model of networks layers, IP protocol, TCP and UDP, network security, and network services (SNMP, HTTP, DNS). It is one of the mandatory subjects they must complete in their degree. The teaching dynamics are based on lectures in the classroom, practical parts in the laboratory, and students' public work presentations to their partners.
Description of students' engagement with Differ and/or Hubert: The educator created an online community in Differ (named "Fundamentos de Redes"), with the help of BO. Students were invited to join the community via internal message (in the official Moodle platform of the University of Granada) and e-mail. It was not mandatory to participate in this community, so just 23 students joined it. The main use of Differ was as a tool to communicate with the educator regarding any question about the subject. Moreover, the students could use it as a breaking-ice utility for contact with their peers. The educator made public announcements in the Differ community and some students used it to ask some questions about the course agenda or doubts about the contents. Most of the students participated at least one time in the chats, mainly asking course management or content questions directly to the educator.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Overall opinion of educators participating in Pilot 2

We could comment here on the feedback of three groups of educators. The first one was those educators participating in the webinar in which Differ tool was presented; where they were encouraged to use it in their subjects. Even if they liked the application and its possibilities, they argued that using it effectively would have some problems, such as the inconvenience for the students of learning to use a completely new tool or the possible issues that could come from the University regarding the use of a non-oficial tool in the class. They also argue that this will imply additional work in the subjects that probably won't be worth the time spent in preparing it. Thus, they didn't see enough advantages in the use of Differ

Second, we could say that educators using Differ in their classroom have a similar opinion to those of Pilot 1. Finally, just one of the educators used Differ in this Pilot (one of the participants in the project), mainly for announcements and communication features. He commented that the students used Differ and interacted with BO, but not as much as it was expected. According to their feedback, they didn't see utility in BO's options, and some of them couldn't deal with the language barrier (Differ and BO just worked in English). Thus, they finally used the community just as an alternative communication tool with the educator.

The third group are those educators who used Telegram in their courses for the complete management of teaching (organisation of groups, direct communication, providing teaching materials, resolving doubts, etc). They argued that Telegram offers functionalities and improvements in university teaching, enhancing communication between lecturers and students. It also facilitates collaborative work, real-time questionnaires, information extension, etc. Moreover, a fundamental advantage is that the students know how the tool works and its features, so it does not require installing and learning a new tool.

Finally, we would like to summarise the opinion that mentors (students working to encourage and help other students joining Differ.chat) participating in this pilot wrote in a set of essays. Most of them complained about technical issues with the application (some interface problems and incompatibilities with some devices), as well as the non-existence of a Linux version, since students are used to working in this Operative System at ETSIIT and in their laptops. There were also some limitations in the chat possibilities that were present in usual communication tools (such as replies or references to other comments in the chat, emojis and stickers, or audio notes). On the other hand, most of them remarked on the positive features of Differ and BO, such as the creation of interesting communities or the ice-breaking capabilities, very useful for students in their first year or Erasmus ones. However, according to the essays, there were two main drawbacks in Differ: the language barrier (it only worked in English during this pilot), and the 'annoying' necessity of installing and learning to use a new application to communicate with their peers, which they normally do already using Telegram or WhatsApp.

3.1.3. Courses at the University of Leeds

Pilot 1 had indicated challenges in the use of Differ and Hubert by educators teaching a course at Leeds University. We offered Differ and Hubert to students on 3 courses: Computing MSc

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

module COMP5840 Data Mining and Text Analytics, class size 136; Computing undergraduate module COMP2121 Data Mining, class size 122; and Computing undergraduate module XJCO2121 Data Mining, taught remotely to 72 students at the Leeds-SWJTU Joint School campus in Chengdu, China. Leeds University IT Service requires students and staff to use other platforms in online communication and issues its own course evaluation surveys. Hence we were unable to make the use of Differ or Hubert mandatory as part of a taught module class, and it was challenging to persuade students to volunteer to trial the use of additional platforms. A handful of students agreed to try Differ and Hubert, but gave up saying they preferred the standard Leeds IT platforms for chat discussion and student feedback.

So, for Pilot 2 we instead piloted the use of Differ as a platform to encourage university student belonging and community in two broader student groups: PhD students in the School of Computing, a cohort of 180 research postgraduate students, mainly International (non-UK) students age over 21; and “mature students”, older students who may feel isolated in classes of mainly younger students. The Lifelong Learning Centre provides support and advice for mature students in all disciplines across the university, mainly UK students aged 21 to over 60.

Educators were not involved directly, as these user groups are not taught classes overseen by an educator. Instead, the student onboarding was handled by administrative staff. The School of Computing director of postgraduate research studies sent invitation emails to all Computing PhD students inviting them to onboard to Differ, using the interface described in previous sections. Lifelong Learning Centre (LLC) professional service staff encouraged mature students supported by LLC to use Differ; they sent announcements to the cohort via email, with URLs to joining web pages. Differ was presented as a chat platform for both cohorts of students (Computing PhD students and campus-wide mature students) to meet, to encourage student belonging and community.

Hubert was not used in Pilot 2, as there was no “taught course” to be evaluated via Hubert. As educators were not involved (the pilot cases were launched by administrative staff), Leeds University academic educators could not fill in a survey on usage of Differ and Hubert on courses.

Best Case

Course name: Mature student community supported by LLC Lifelong Learning Centre
Number of students: 3650+ students were invited; of these, nearly 300 joined (less than 10%), c9000 messages sent via Differ.
Course description: Pilot 1 had indicated challenges in use of Differ and Hubert by educators in teaching a course; so instead we piloted use of Differ as a platform to encourage university student belonging and community in a broader student group: “mature students”, older students who may feel isolated in classes of mainly younger students. The Lifelong Learning Centre provides support and advice for mature students in all disciplines across the university, mainly UK students aged 21 to over 60.
Description of students’ engagement with Differ: Lifelong Learning Centre (LLC) professional service staff encouraged mature students supported by LLC to use Differ but it was not compulsory. They sent announcements to the cohort via email, with URLs to joining web pages.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Differ was presented as a chat platform for students to meet, to encourage LLC student belonging and community. About 10% of the students joined Differ and a significant minority used Differ to communicate with other student(s): c. 9000 messages were sent via Differ.
Affordances and challenges: Leeds University IT Service requires all students and staff to use other chat platforms in online communication, and we could not make use of Differ mandatory. It was challenging to persuade students to volunteer trial use of this additional platform; however a significant minority of LLC-supported mature students liked the platform and continued to use it. The methods and results were presented in "Online Chat and Chatbots to Enhance Mature Student Engagement in Higher Education" submitted to International Journal of Lifelong Education; and "Edubots: chatbots for university student belonging and community" submitted to the Student Education Conference, 6-7 Jan 2022, Leeds University.

Overall opinions and conclusions

Mature students transitioning into their first year of higher education face many difficulties that affect their motivation, participation and success. Their feelings of being disconnected from their peers and from their institutions, in addition to the challenges they face while managing their finances and family responsibilities, are among the key barriers to the successful completion of their courses. Encouraging online student engagement among mature students to establish social connection with their instructors and peers can reduce their isolation and enhance their 'sense of belonging'. During the academic year 2020/2021, the Lifelong Learning Centre (LLC) at the University of Leeds in the UK, decided to pilot an online chat platform 'Differ' including a chatbot 'Bo' after it had seen a decline in the use of programme Facebook groups. During the pilot, feedback was collected from students. A mixed-methods approach was adopted to boost robustness through triangulation. Several forms of data collection methods were used: a mentimeter, an online survey, three focus group sessions and a semi-structured interview. This study sheds light on the different aspects of creating student-led online communities and provides recommendations on how to improve the uptake of students in the future.

In this study, using Differ to assist students with social integration had a positive impact among mature students by facilitating the formation of initial connections with their peers. This enhanced students' sense of belonging and engagement with other LLC students. The survey responses from students who went on to use Differ show that 59% actively engaged with Differ, 63% found Differ useful and 59% would recommend Differ to a friend. Differ may be particularly appropriate to boost student engagement among non-standard student cohorts, for example International students joining the Summer School, or students enrolled in interdisciplinary programs. We recommend setting norms and rules for these communities to maintain safe spaces for students to communicate. Moreover, having a separate tool for university student use only can help students draw the line between their student life and their private life.

To improve the uptake of Differ in the future, applicants can be invited to join Differ in the pre-entry academic skills programme. This will provide a method of keeping in touch, and connecting with peers, rather than waiting until September/October arrival at Leeds. In addition, as some students experienced confusion over which communities to join and missed the email invitations,

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

students can be invited using SMS which should make their joining experience smoother. Also, the LLC professional services staff can design a training programme for new students that provides instruction on how to create new communities and on how to use the 'ask a question' feature to find peers, to make it easier for students to navigate Differ. Furthermore, feedback indicates that communicating via Differ has encouraged more formal language use which has influenced interactions with students at times; future training could encourage them to be less formal and to have more fun with their interactions.

Understanding the educational and social experiences of mature students can enable universities to shape their culture and structure to suit the needs and interests of these students, particularly those transitioning into their first year of their courses. In this study, feedback indicated that, before starting the university, students felt more confident they would be supported during their study. This support is vital to mitigate many of the challenges faced by minority-group students and to improve their retention in HE through providing assistance with social and academic integration. For instance, staff can assist by integrating information about services and support into teaching. Moreover, institutions should integrate chat tools that can be utilised by both educators and students to their larger educational systems. This can reduce the unnecessary stress of learning different tools and promote the use of these tools among students.

3.1.4. Courses at the University of Zagreb

Educators from the University of Zagreb were invited to explore the use of chat tool **Differ** through the University's web pages (see Figure 24. Invitation to educators at University of Zagreb and Figure 25).

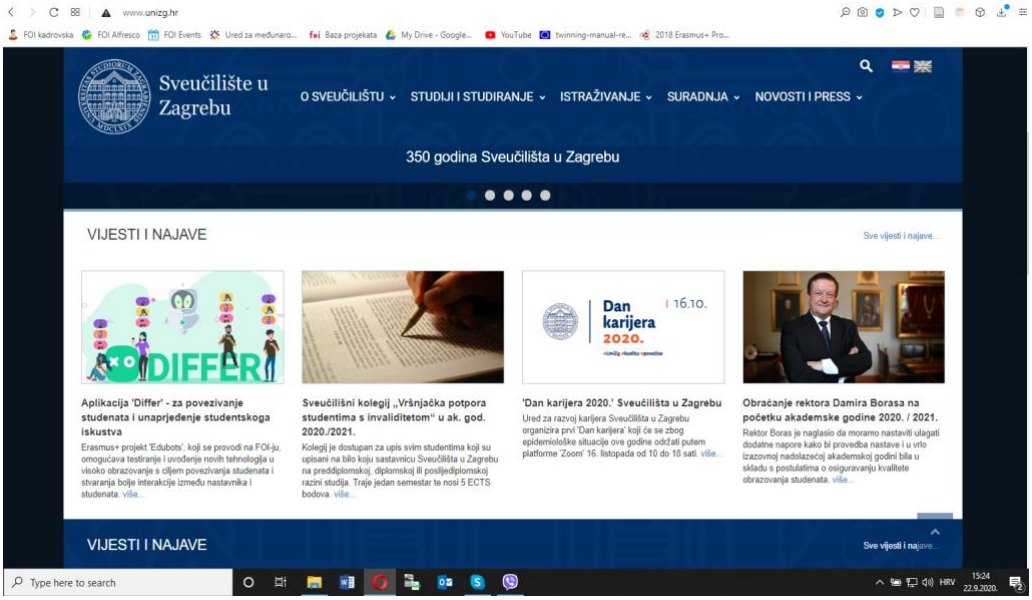


Figure 24. Invitation to educators at University of Zagreb

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

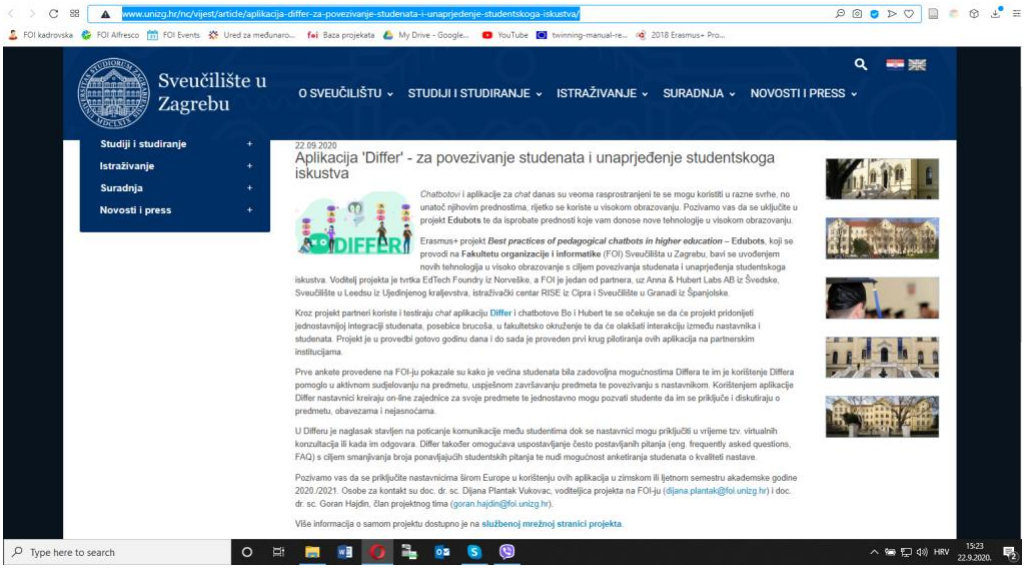


Figure 25. Invitation to educators at University of Zagreb

Since no response from educators was received by that channel, an invitation was sent to the community of the teaching assistants at the Faculty of Organization and Informatics. For them, a webinar was organized on which characteristics of Differ were explained and good practices about using Differ for teaching purposes from previous courses were introduced (see Figure 26-Figure 29).

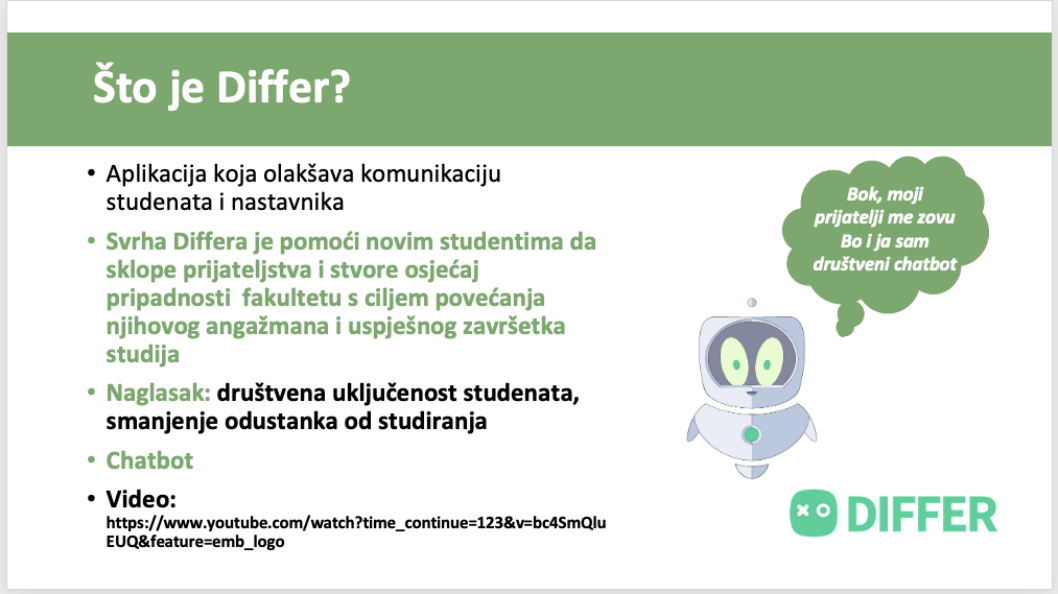


Figure 26. Slides from onboarding presentation about Differ and BO

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Figure 27. Slides from onboarding presentation about Differ and BO

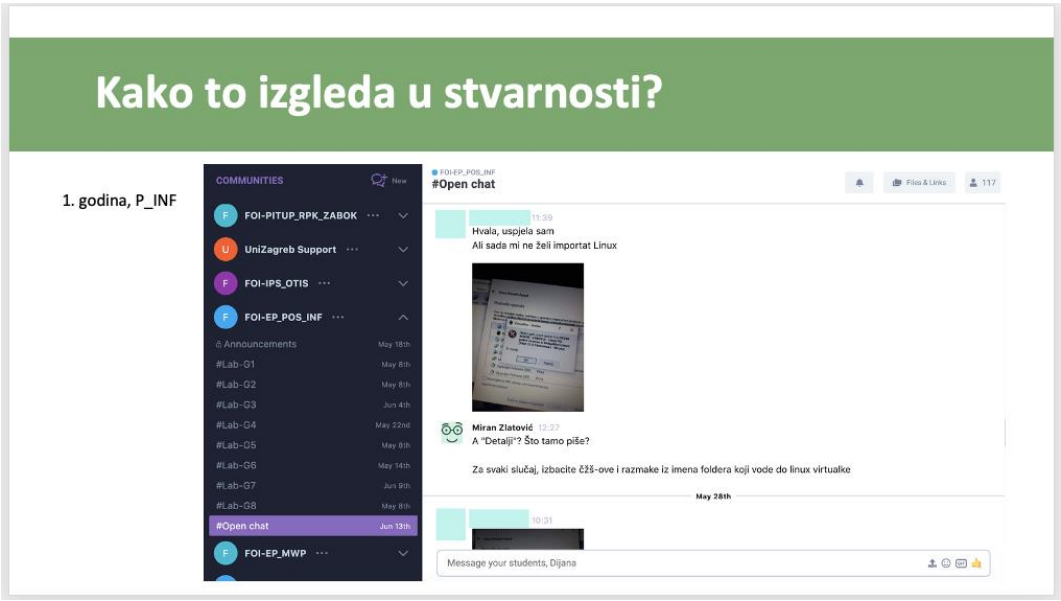


Figure 28. Slides from onboarding presentation about Differ and BO

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

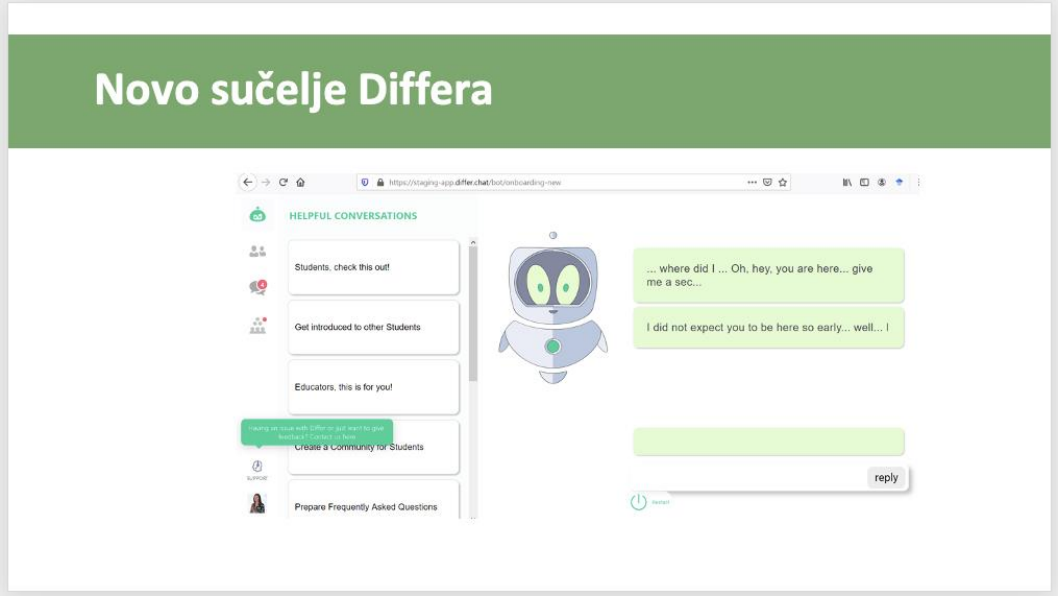


Figure 29. Slides from onboarding presentation about Differ and BO

Interested educators also participated in the webinars organized by EdTech in which new functionalities of Differ were introduced, particularly the communication with the chatbot BO. To all interested educators guidelines with the procedure about how to start with Differ was sent by an email (see Figure 30), together with a User manual in Croatian that was created during Pilot 1.

Poštovane kolegice i kolege,
hvala Vam na iskazanom interesu za korištenjem aplikacije Differ u nastavi i komunikaciji sa studentima. U nastavku šaljem opis početnih koraka u radu s aplikacijom.

Prvi koraci u Differu

1. molimo ispunite prijavu na https://join.edubots.eu/sign-up
2. instalirajte aplikaciju Differ (mobile/desktop + korištenje web sučelja)
3. isprobajte chatbota Bo-a i pomoću njega instalirajte zajednicu za svoj predmet (ili mi javite pa će to odraditi FOI tim i poslati vam join link za vaš predmet)
   - preporuka za naziv zajednice: naziv fakulteta i skraćeni naziv predmeta, npr. FOI-MMS-PITUP
4. pridružite se zajednici (community) **Edubots Educators**
   https://join.differ.chat/community/5b0a193b-4829-4d2a-a5f2-0907826296e5/LTZRKH
   (predstavite se u #SayHi)

Pomoć FOI tima

- kreiranje zajednice za predmet (FOI tim ili Differ admin)
- pridružite se zajednici **UniZagreb Support** - tu na hrvatskom jeziku razmjenjujemo savjete i primjere dobre prakse za nastavnike i studente demostratore. Poveznica na zajednicu:

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

https://join.differ.chat/community/7aa7ba8e-341c-4aa6-af7d-73684e500265/IMQQQV

- u zajednicu UniZagreb Support poslat ćemo primjere pozivnih poruka za Moodle i pozdravnih poruka na Differu - pošaljite ih na početku svoje nastave, npr. na prvim-drugim seminarima ili lab vježbama
- **u privitku su upute** o korištenju Differa, na hrvatskom jeziku
- online upute na engleskom jeziku nalaze se na Differ Knowledge Base

Prijedlog organizacije predmeta u Differu

- prema temama #topics (osnovne su Announcements, Open chat, a ostale sami izrađujete)
- struktura tema prema grupama (npr. G1-pon 12-14h) ili temama vježbi/seminara
- angažirajte studente-mentore ili demonstratore da čim više odgovaraju na studentske poruke
- u Announcements objavite sate virtualnih konzultacija kada ćete biti dostupni live za odgovaranje na poruke
- potičite studente da sami diskutiraju i predlažu rješenja

Ljubazno molimo da kreirate zajednicu za svoj/e predmet/e (ili javite da mi to učinimo) do **10.10.** i pozovete svoje studente da se pridruže zajednici. Primjeri poziva bit će stavljeni u zajednicu **UniZagreb Support.**

Stojimo na raspolaganju za sva dodatna pitanja.

Hvala još jednom i lijep pozdrav,
FOI tim Ana, Antonela, Goran, Didi i Dijana

Figure 30. An email messages with the guidelines for Croatian educators

Altogether, Differ was used on **eleven courses** from the University of Zagreb, of which 10 courses were from the Faculty of Organization and Informatics, while one course was from the Faculty of Electrical Engineering and Computing.

Chatbot **Hubert** as a tool for course evaluation was also used on these eleven courses. By the end of the semester, educators were provided with the guidelines on how to create a course evaluation with Hubert (or a project administrator created an evaluation for the educator). Guidelines were created in the English language (see Figure 31-Figure 36) and disseminated to all EDUBOTS partners.

EDUBOTS WP5, T5.6 Continuous knowledge sharing, Pilot 2

Instructions for creating Hubert evaluations

Each partner should prepare an environment for implementation of the surveys. We suggest the following:

- Create and access your Hubert account, https://hubert.ai/signup
- Follow the steps in creating Hubert evaluations

If you would like to test Hubert first, please create the test evaluation first (step 2: insert the „Test" prefix into the evaluation's name) and later create production evalutation for the course.

You could also use this test evaluations, FOI test 2 EDUBOTS Pilot 2:
https://www.hubert.ai/feedback/5fd391d07dcd1f0011d815ef

Figure 31. Hubert guidelines in English language

Co-funded by the Erasmus+ Programme of the European Union

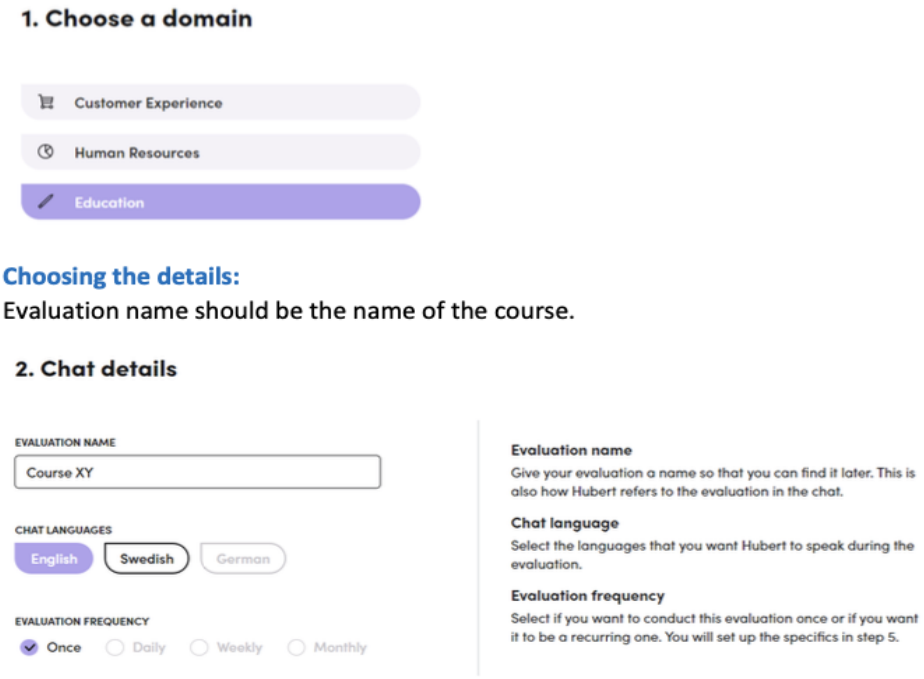
THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Steps in creating Hubert evaluation

After login, from the main screen (Dashboard) choose *Create a new evaluation* → *Education*.
Other navigation is possible: from the dropdown menu *Evaluations* choose *Evaluations* and then the button Create new.

There are 5 steps in creating the survey:

1. **Choosing the domain Education:**



2. **Choosing the details:**
Evaluation name should be the name of the course.

Figure 32. Hubert guidelines in English language

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

3. **Choosing the chat template:**
   Please choose the template **EDUBOTS Pilot 2.**

   If you do not see the EDUBOTS Pilot 2 template, please contact Viktor Nordmark viktor@hubert.ai or Jacob Mattsson jacob@hubert.ai to connect your account to the template.

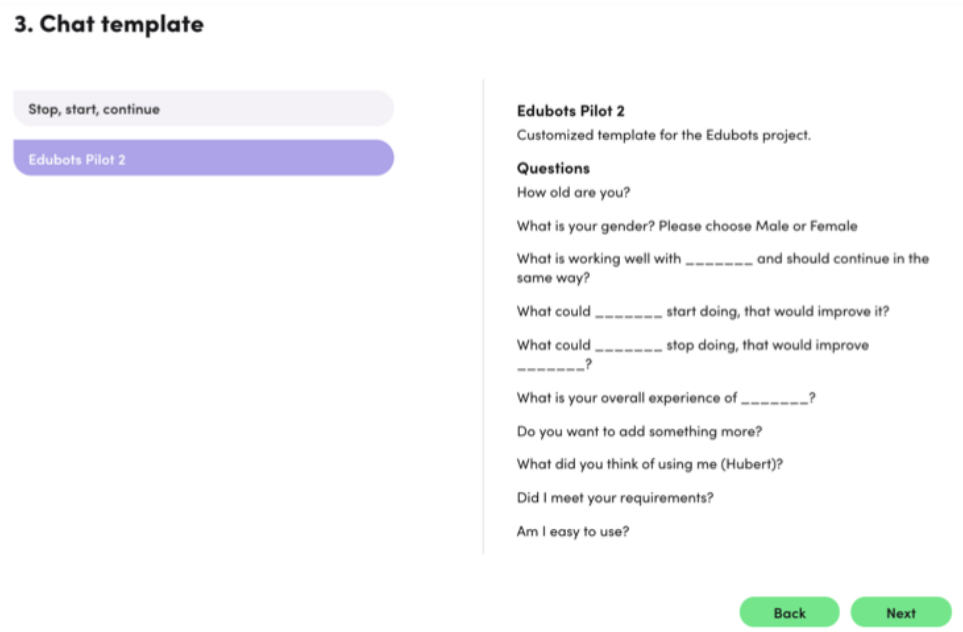


Figure 33. Hubert guidelines in English language

4. **Additional information:**

   (suggestion for the invitation message at the beginning of the survey)

   Hello, here is a short survey about your satisfaction with the course (insert the course_name). We kindly ask you to answer the questions honestly in English language. Chatbot will guide you through the survey, and don't worry - it is completely anonymous. Now, you know everything, let's talk. :-)

   How should Hubert call you?
   * teachers (or insert the teacher's name if you would like to evaluate the teacher instead of the course)

   What are you evaluating?
   * (insert the course_name)

Figure 34. Hubert guidelines in English language

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

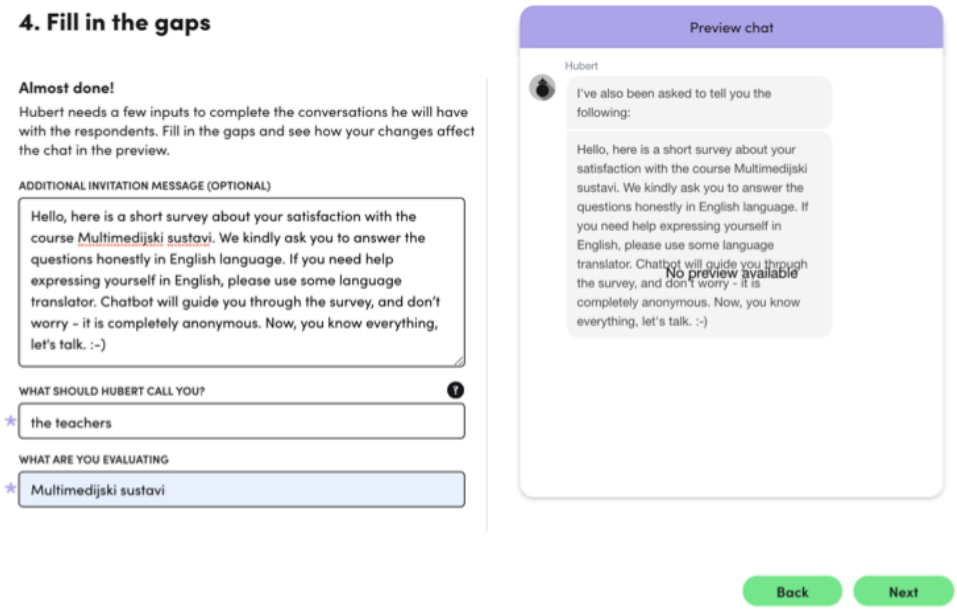
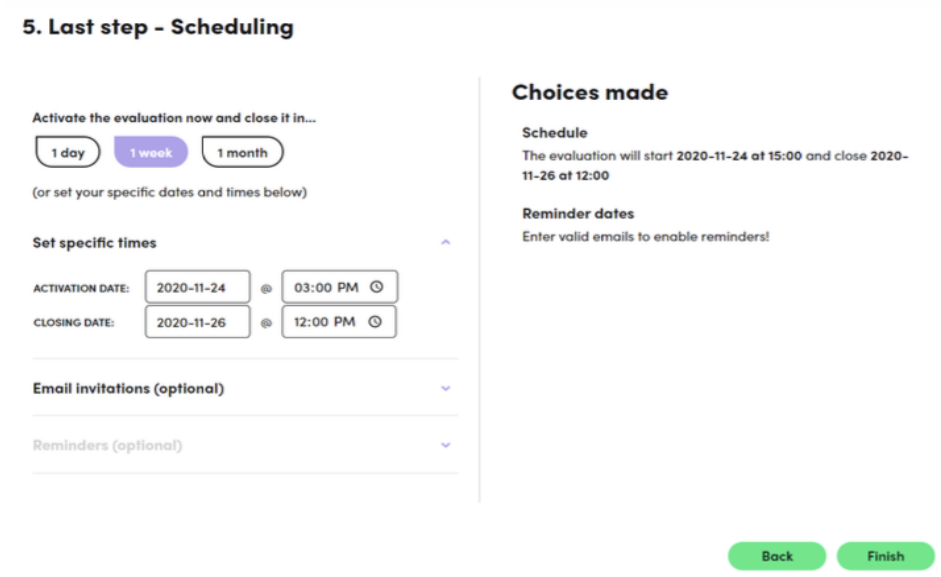


Figure 35. Hubert guidelines in English language

5. **Finishing the evaluation:**
Choose the options about activation of the



After this step you can generate the link for sharing the survey with your students.

Figure 36. Hubert guidelines in English language

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Best Case Study from the University of Zagreb

Course name: Business Communication (Faculty of Organization and Informatics)
Number of students: over 100
Course description: The course is held in the 1st year of bachelor's studies. Teaching activities were fully online except the first and the third week when seminars were held face to face. Goal of the Business Communication course is to teach and prepare students for various forms of business interactions like presentations, negotiation, sales, written communication, and writing CV-s. Students were mostly 18-19 years old full time students.
Description of students' engagement with Differ and/or Hubert: Educators on the course used chatbot Bo within Differ to create the community. Primary focus of Differ use on the course was to get introduced to other students (ice-breaking), ask questions to the community, and creation of the community for students. Differ helped students to connect with their peers especially for team-work presentation tasks, but also with student mentors who helped them in preparation for presentation. Students often used Differ before exams to ask the educator some additional questions, or after the exam. In the cases when they failed the test, Differ was a good tool to encourage them for further learning and explain how to fix bad results later in the semester. They often feel happy and relieved because they have quick information that helps them a lot to stay motivated for the course. Differ helped educators get closer to students which was very meaningful since lectures were almost fully online. Student-mentor used Differ to prepare students for their projects, group and peer-to-peer-communication.
For each study group (approximately 40 students per group), a separate community was created which is shown on Figure 37. For the whole course, eight (8) Differ communities were used for communication with students during the semester. This approach was convenient to bond students who were preparing the presentations in small groups and had to solve practical exercises during the lectures throughout the semester. Because of this, they were encouraged by the course educator to use Differ for such learning and social activities. From the educator perspective, in Differ it was easier to manage smaller groups then one larger (approximately 300+ students). The main argument is that online Differ communities were the equivalent for weekly exercise lectures in which every group had different presentation topics.
Affordances and challenges: Main affordance is that communication was faster, informal, but yet academic appropriate. Challenges were primarily seen in students who are not willing to try out something new because they already use different communication tools. They were overloaded with everything online due to the COVID-19 situation.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

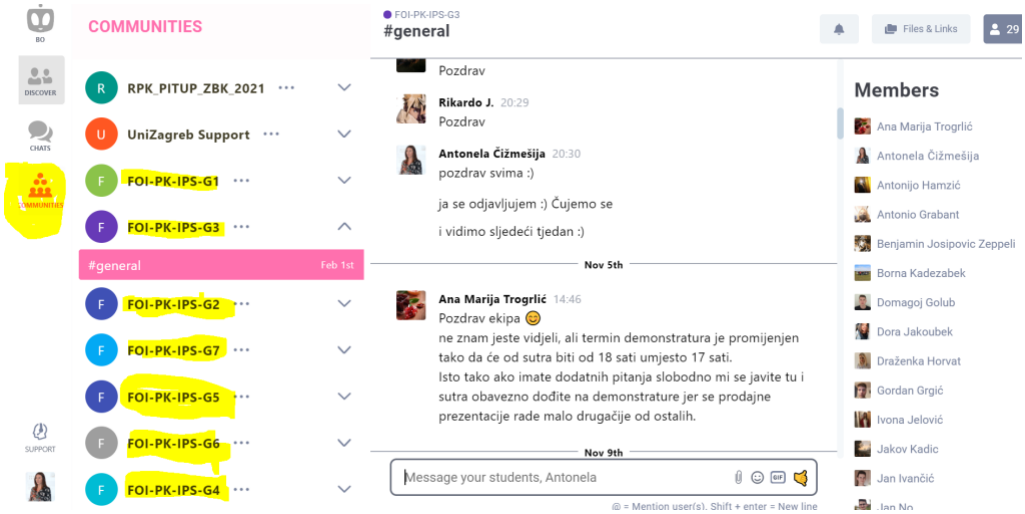


Figure 37. Different communities in Differ for the same course in Pilot 2.

Overall opinion of educators participating in Pilot 2

We have gathered opinions of three educators that were using Differ in Pilot 2 courses. All of them expressed a positive opinion about Differ in regards to enhanced and faster communication with students. The communication was less formal but appropriate one. Using Differ helped them to connect with the students and encouraged their social participation in the course. They all expressed that they would like to use Differ at their other courses. However, two of three educators had neutral opinions about Differ as a teaching tool: they neither agree nor disagree that Differ helped them to share their knowledge with students.

Hubert was used to collect students' feedback about the course. Evaluation was conducted at the end of the semester, after all course activities ended and grades were known and students were not interested in the subject. Because of this about 20 students gave an opinion on Hubert. From a educator's point of view, collected data from open-ended questions were very useful because they complemented students' ratings of educators. Because of COVID-19 situation and dominantly full online teaching during the semester where little information (feedback) could be collected from students, this kind course evaluation was meaningful to reconsider teaching strategies in future.

3.2. Interactions in Differ and Hubert

3.2.1. Analysis of system data in Differ

The Differ system data collected during Pilot 2 and the community categorisation were as described in section 2.4.1 Analysis of system data in Differ above. From Pilot 2 onwards, when

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

the BO chatbot was made available, additional events relating to users' interactions with BO were also captured.

The number of communities created and volume of users ever active in Pilot 2 are shown below.

Community Type	Pilot 2 Communities	Users Ever Active by Differ Community		
		Minimum	Maximum	Mean
Course	15	0	63	19.6
Programme	16	1	36	12.6
Department	1	28	28	28.0
Multi-Disciplinary	2	2	87	44.5
Total	**34**	**0**	**87**	**18.0**

Table 5. Number of communities and active users in pilot 2

3.2.2. Analysis of data from the course evaluations in Hubert

See section 5.2.2. Analysis of data from the course evaluations in Hubert below for analysis of Hubert activity across all pilot phases.

3.3. Results of the surveys and interviews about the chatbots

3.3.1. Differ and learning process

In Pilot 2, survey about Differ and learning process and it's User Experience and Usability was conducted on the Universities that are part of the EDUBOTS team. The goal was to find out how students used Differ as well as how it helped them in the learning process. All collected data was anonymous and voluntary. Data presented in this section represents summarised research findings from University of Zagreb, Faculty of Organization and Informatics, University of Granada and CYENS (former RISE) **who agreed on using the same questionnaire.** University of Leeds participated with 27 students, but their questionnaire **differentiated** from other partners and data analysis from this questionnaire is not presented here. Research was conducted in the winter semester of the academic year 2020./2021. In total, 88 responses were collected from the students. After cleaning the data and excluding student's responses who did not use Differ in courses, **the sample consisted of 71 participants, N=71.** By home country, the participant's structure is presented on Figure 38.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

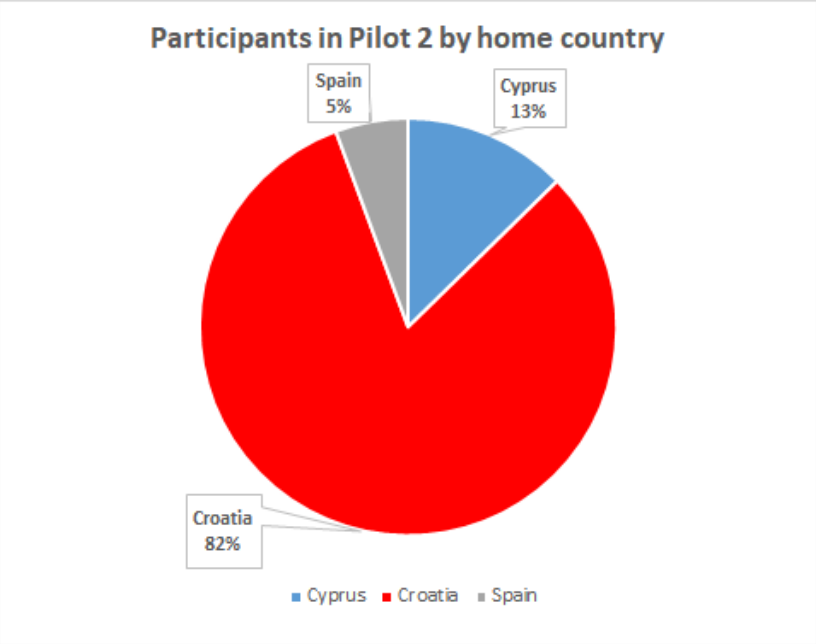


Figure 38. Participants in Pilot 2, N=71

22 female students, 48 male filled out the survey. One of the students preferred not to say gender. By age structure, most of the students were 19-22 years old and only 4 had 30 years or more.

Regarding frequency of Differ use, results are shown on Figure 39. Most of the students used Differ several times a month or once a month (approximately 68% of respondents) and a minority of students have used it on a regular basis (weekly or daily).

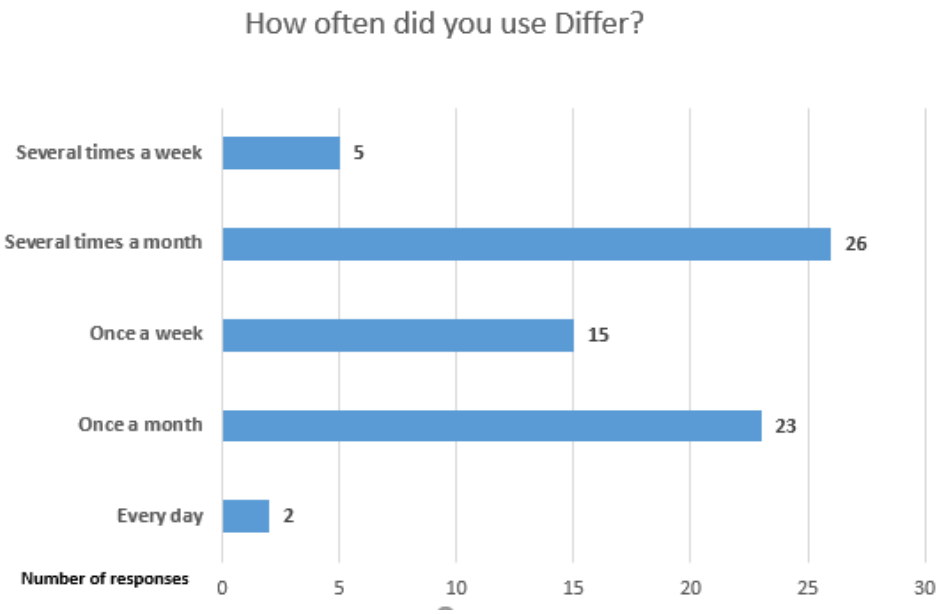


Figure 39. Frequency of Differ use in Pilot 2

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Most of the students (81,9%) used mobile applications to access course communities in Differ (Figure 40).

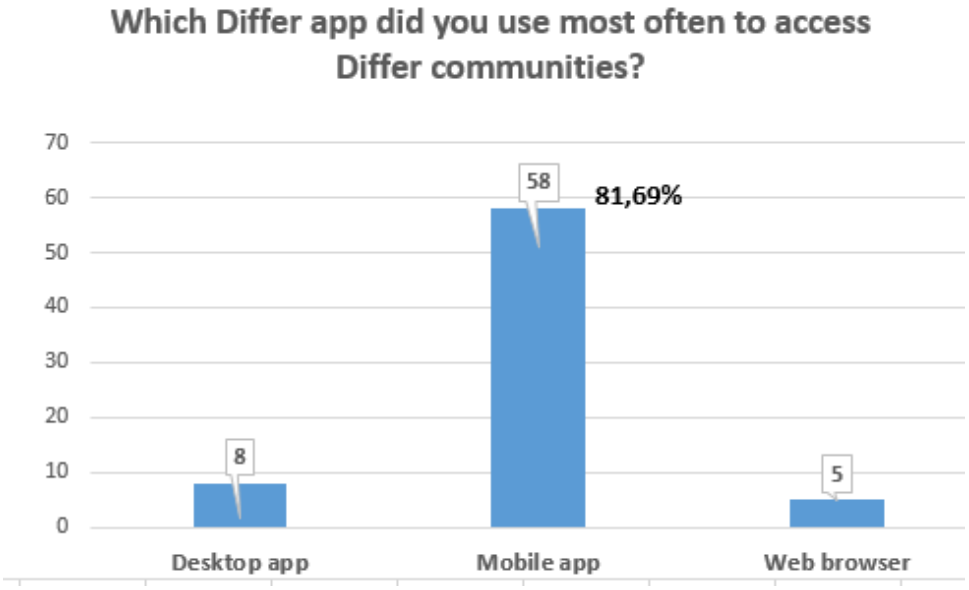


Figure 40. Applications to access Differ communities

The following graph on Figure presents distribution of answers on a five-point scale (from strongly disagree to strongly agree) related to the 10 statements. In total, **62 respondents** from primarily 71 filled out this part of the survey related to the role of Differ in the learning process. While the dominant answer for most statements is neutral, students felt free to express themselves through Differ and connect with course educators. On the other hand, a lot of students did not think that Differ helped them connect with other students nor did it help them share their knowledge. We can speculate that students retained to use previous ways of communication, like Facebook, email, phone calls and other ways of communication they were familiar with before they were introduced to the Differ. Also due to COVID-19 outbreak students were oversaturated with different new tools educators used to enhance online lectures.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

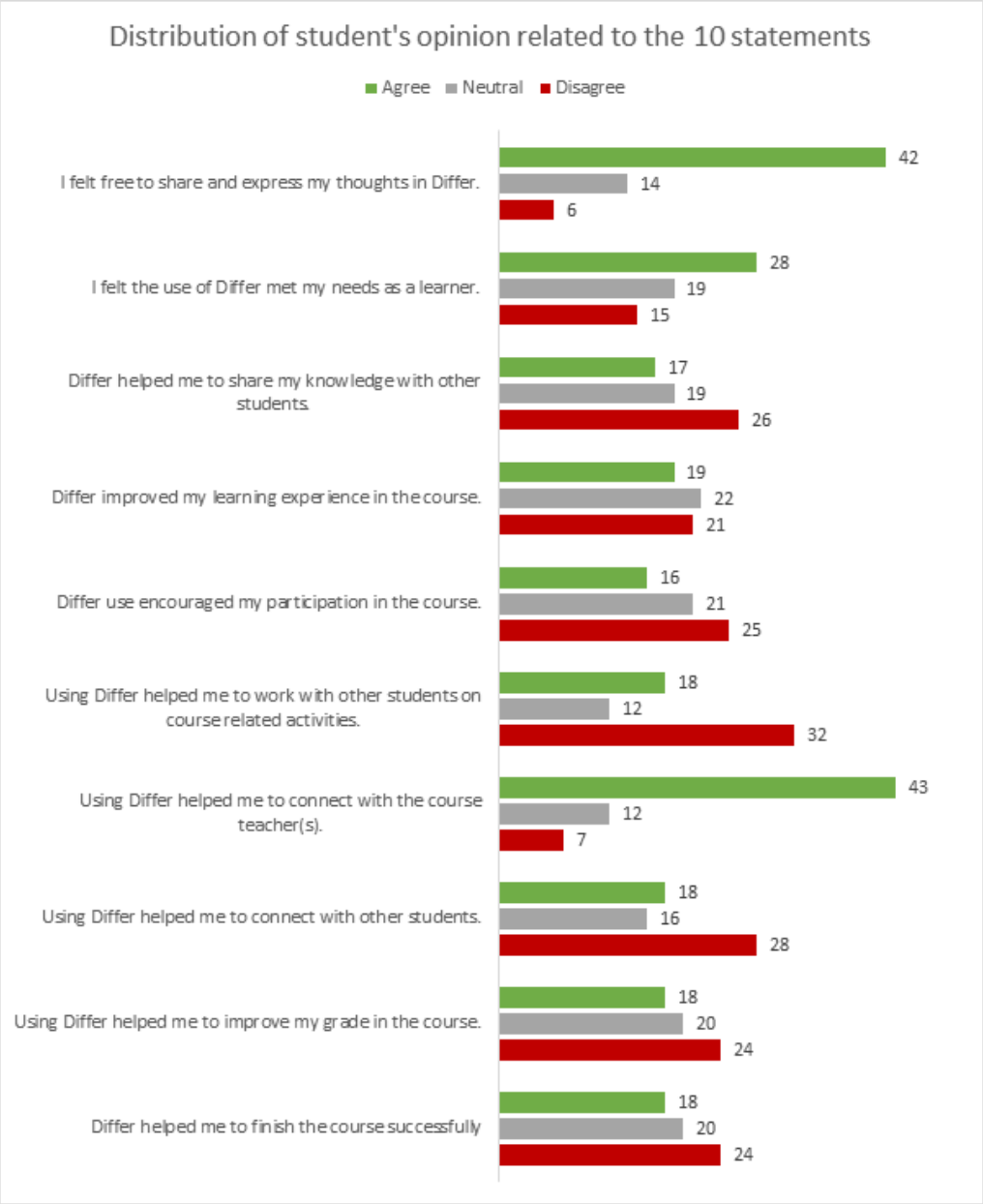


Figure 41. Distribution of students' opinions related to the 10 statements, N=62

Comparison of Students' and Educators' Survey About Differ Use

The conducted survey does not have a representative sample and thus the results should not be considered as scientific conclusions, the results indicate more **positive attitude towards the Differ use among the educators**. On the other hand, **students have a more negative attitude about the Differ** use When comparing average results, the distribution of educators' answers is 24% negative, 19% neutral and 57% positive, while students' answers are on average distributed 34% negative, 28% neutral and 38% positive. A large **difference is present in the statement**

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

that Differ helped a person connect with other students. From a student's perspective that is false in over 40%, while from the educator's perspective that is false in less than 15%. On the other hand, less than 15% of students also perceived that Differ did not help them connect with a course educator. This indicates that **students connect with other students via other tools** and ways of communication, while both educators and students perceive Differ as a potent tool to connect between the groups.

On the other hand, over 55% of educators perceived that Differ did not help them improve their teaching, while less than 40% of students perceived that Differ did not improve their learning experience. Contrary to the communication aspect, it can be seen that **students perceive Differ as a good support for their learning process, while educators do not think it improves their teaching**.

Both groups agree that **Differ did not help them share the knowledge**. They also agree that **Differ made them feel free to share their thoughts** in the application. From those two statements it could be concluded that Differ was used more as a social application than a learning tool.

3.3.2. Differ and teaching process

Educators' Survey About Differ and Hubert Use

In the Project as the part of Pilot 2 phase we performed a survey about use of Differ and Hubert with educators. The goal was to find out how educators use Differ and Hubert as well as how it helps them in the teaching process. Only 8 educators participated in the survey so all quantitative data will not be further interpreted, but just presented as is.

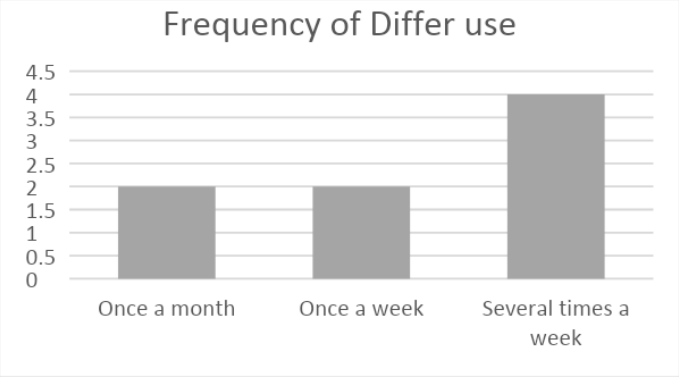


Figure 42. Frequency of Differ use

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

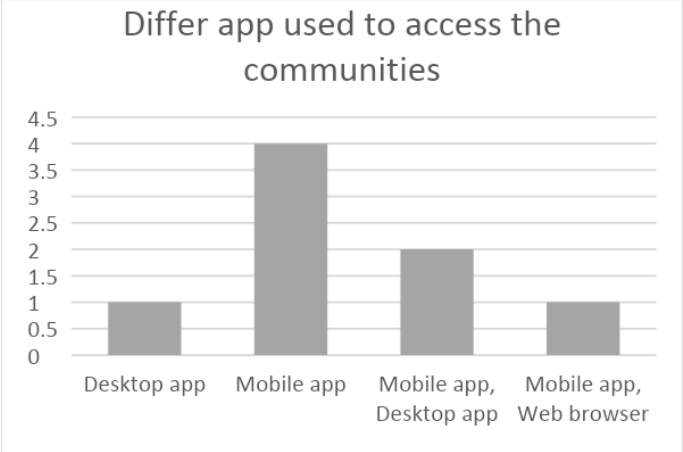


Figure 43. Differ app used to access the communities

Three educators mostly used desktop while five educators mostly used mobile apps.

Two educators used Hubert on 3 courses, one on 2, one on 1 while four did not use Hubert. Educators used Hubert for course evaluation only once in the semester. Educators who used Hubert thought that it had a positive impact on the proportion of students successfully completing the course.

Dominant way of joining Differ was word of mouth – through an invite from a colleague. One educator joined through the faculty's webinar, one by the university's website and one by invitation on the EDUBOTS webpage.

Dominant way of creating a course community in Differ was interaction with the Bo – the chatbot.

Most courses were presented in the form of lectures and laboratory exercises, while few had lectures and student seminars. Most of the courses were enrolled by students which were around 22 years old, while one course had younger students which were dominant 18 and 19 years old and another course which was focused on migrants which ranged from 20 to 35 years.

The following two graphs present educator's answers related to the use of Differ in the teaching process. Since the number of collected answers is low all interpretation should not be considered scientific conclusions, but more in line with preliminary indications. Both graphs present a number of statements which were evaluated by educators on the five-point scale (from strongly disagree to strongly agree).

First graph focuses on the **educator's perception of Differ's impact on students**. Results indicate dominant positive attitudes in all statements except for the inappropriate use of communication style, where educators dominantly disagreed, thus their perception is that students used appropriate communication style.

Second graph focuses on nine statements related to the **Differ's impact on the teaching process**. In most of the statements there is a clear positive stance among educators. Two statements which stand out in the opposite direction are related to the improvement of teaching

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

experience in the course and better student engagement in the course due to the Differ usage. In those two statements educators dominantly disagreed.

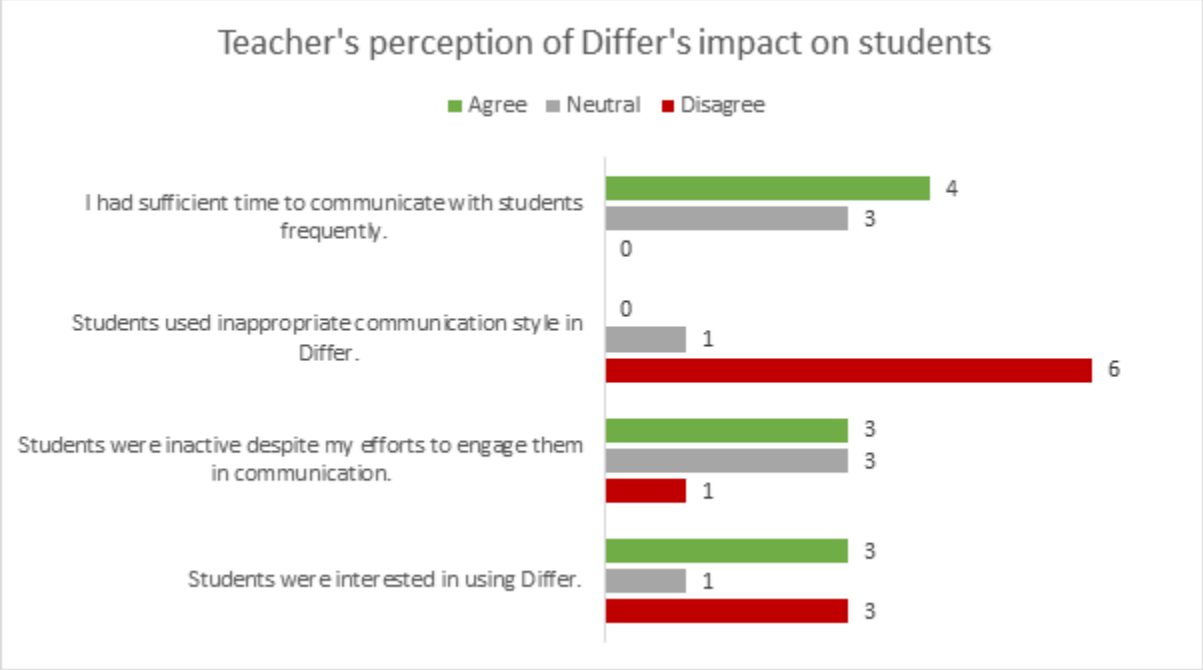


Figure 44. Educators' perception of Differ's impact on students

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

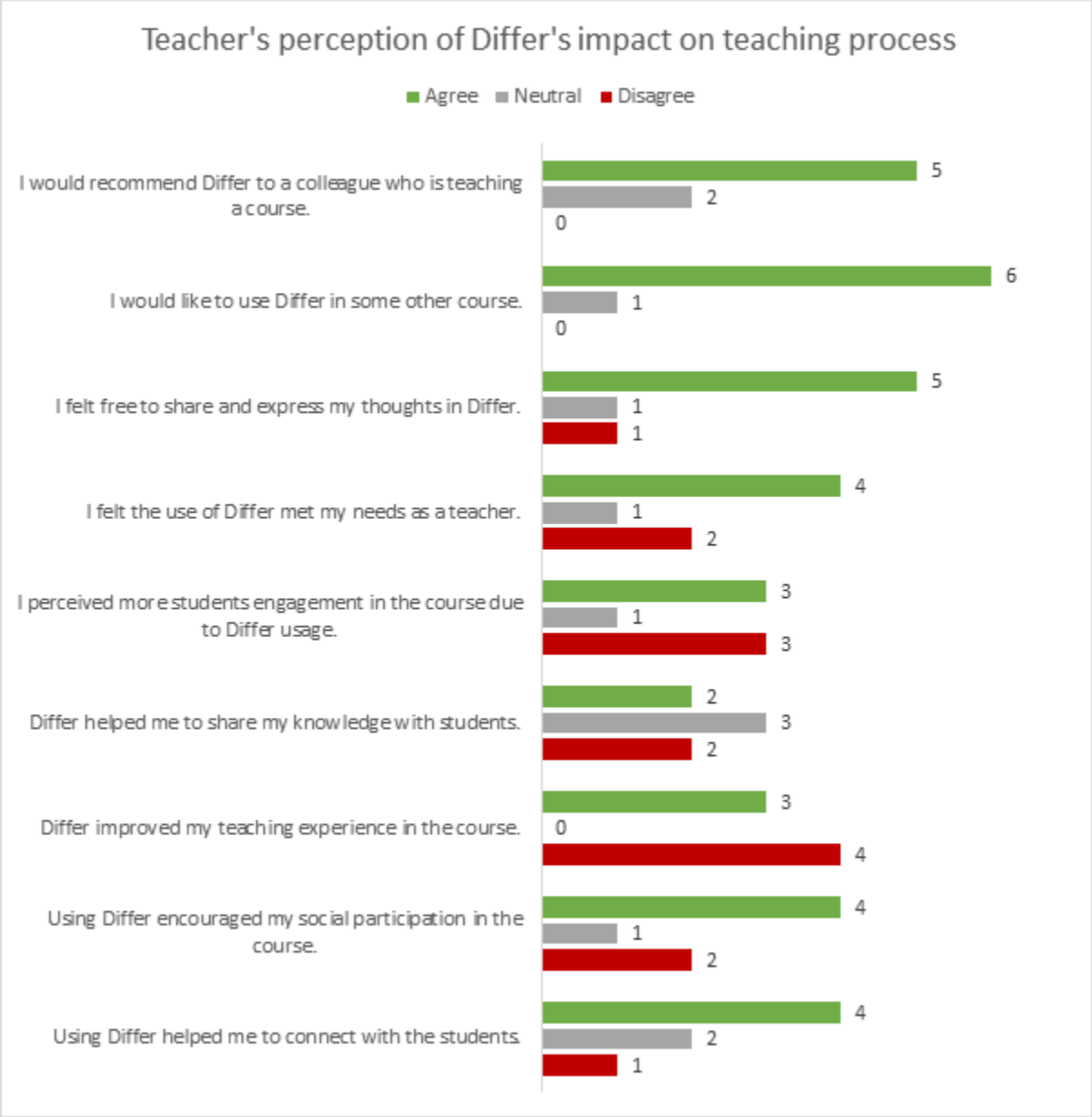


Figure 45. Educators' perception of Differ's impact on students

Comparison of Students' and Educators' Survey About Differ Use

The conducted survey does not have a representative sample and thus the results should not be considered as scientific conclusions, the results indicate more **positive attitude towards the Differ use among the educators**. On the other hand, **students have a more negative attitude about the Differ** use When comparing average results, the distribution of educators' answers is 24% negative, 19% neutral and 57% positive, while students' answers are on average distributed 34% negative, 28% neutral and 38% positive. A large **difference is present in the statement**

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

that Differ helped a person connect with other students. From a student's perspective that is false in over 40%, while from the educator's perspective that is false in less than 15%. On the other hand, less than 15% of students also perceived that Differ did not help them connect with a course educator. This indicates that **students connect with other students via other tools** and ways of communication, while both educators and students perceive Differ as a potent tool to connect between the groups.

On the other hand, over 55% of educators perceived that Differ did not help them improve their teaching, while less than 40% of students perceived that Differ did not improve their learning experience. Contrary to the communication aspect, it can be seen that **students perceive Differ as a good support for their learning process, while educators do not think it improves their teaching**.

Both groups agree that **Differ did not help them share the knowledge**. They also agree that **Differ made them feel free to share their thoughts** in the application. From those two statements it could be concluded that Differ was used more as a social application than a learning tool.

3.3.3. Usability and User Experience of Differ

As in the Pilot 1, a SUS questionnaire (Bangor et al., 2008) was implemented in Pilot 2 courses to collect opinions from the users, primarily students, about usability of Differ. User Experience Questionnaire (UEQ) was used to measure usability and hedonic quality aspects of Differ by applying 26 pairs of contrasting attributes (Schrepp et al., 2017; User Experience Questionnaire (UEQ), n.d.).

Differ Usability

In comparison to the Pilot 1 results, usability of Differ was scored a bit lower in Pilot 2, with 71 participants who evaluated usability (see Table 6).

Pilot 1	N=76		**Pilot 2**	N=71
Mean SUS score	**72,4**		Mean SUS score	**66,8**
StDev	0,80		StDev	0,84

Table 6. Values of Differ SUS scores in Pilot 1 and Pilot 2

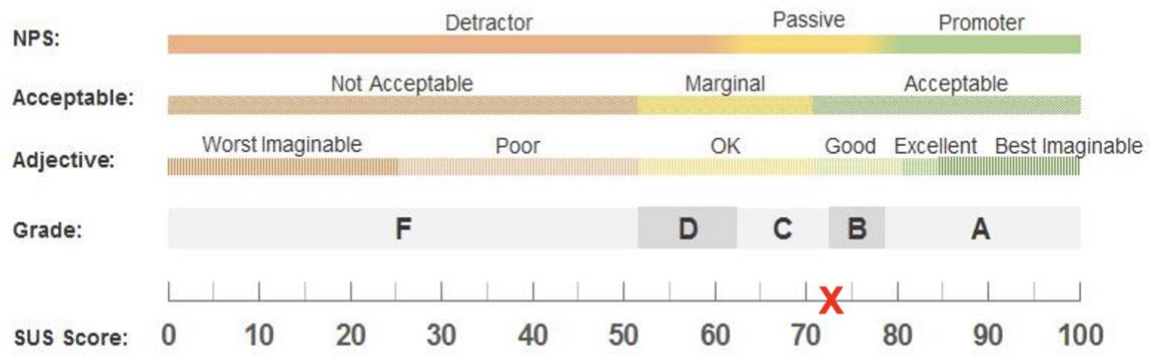
A SUS score above a 68 would be considered above average and anything below 68 is below average.

Other scales that are used to express the usability scores are shown on Figure 46.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Pilot 1



Pilot 2

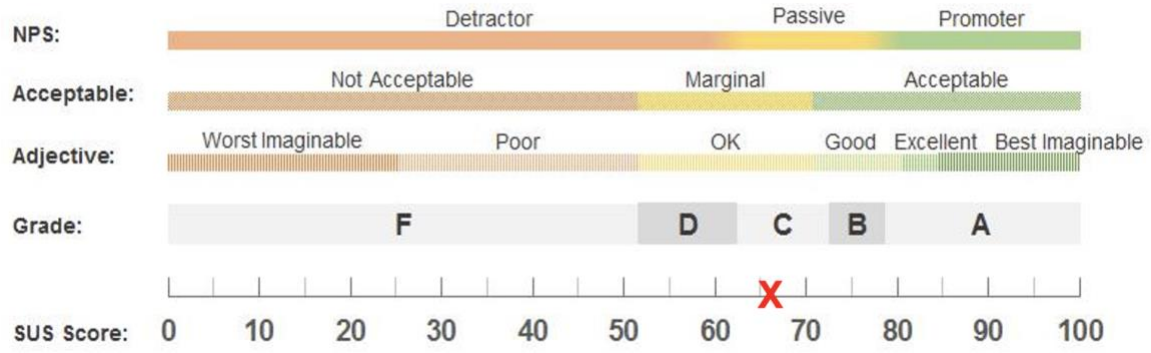


Figure 46. Usability of Differ expressed in SUS score, adjective rating, school grade, interface acceptability and Net Promoter Score

Differ User Experience

User Experience of Differ was measured with the User Experience Questionnaire (UEQ). It is a fast and reliable questionnaire to measure the User Experience of interactive products. Both classical usability aspects (efficiency, perspicuity, dependability) and user experience aspects (originality, stimulation) are measured. The questionnaire consists of 26 pairs of contrasting attributes that may apply to the product. That 26 items relates to 6 scales (Schrepp et al., 2017):

1. **Attractiveness** - Overall impression of the product. Do users like or dislike the product?
2. **Perspicuity** - Is it easy to get familiar with the product? Is it easy to learn how to use a product?
3. **Efficiency** - Can users solve their tasks without necessary effort?
4. **Dependability** - Does the user feel in control of the interaction?

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

5. **Stimulation** - Is it exciting and motivation to use the product?
6. **Novelty** - Is the product innovative and creative? Does the product catch the interest of users?

Results

Values between **-0.8and 0.8 represent a more or less neural evaluation** of the corresponding scale, values > 0,8 represent a positive evaluation and values < -0,8 represent a negative evaluation.

The range of the scales is between -3 (horribly bad) and +3 (extremely good). But in real applications, in general, only values in a restricted range will be observed. It is due to the calculation of means over a range of different persons with different opinions and answer tendencies (for example the avoidance of extreme answer categories) extremely unlikely to observe values above +2 or below -2.

In comparison to the Pilot 1 results, students in Pilot 2 have perceived Differ better in the aspect of hedonic quality, namely stimulation and novelty, as seen in Table 7 and Figure 1. Attractiveness has also gained slightly better value. We could assume that this is related to the introduction of BO, a chatbot within Differ that replaced a Buddy Bot. However, a pragmatic quality has declined according to the responses from the students.

Pilot 1 (N=76)		
UEQ Scales (Mean and Variance)		
Attractiveness	1,024	1,92
Perspicuity	1,789	1,69
Efficiency	1,240	2,14
Dependability	1,260	1,00
Stimulation	0,438	1,67
Novelty	0,572	1,96

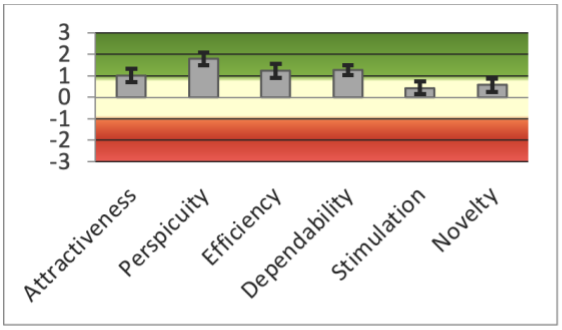
Pilot 2 (N=71)		
UEQ Scales (Mean and Variance)		
Attractiveness	1,190	1,58
Perspicuity	1,581	1,28
Efficiency	1,292	1,42
Dependability	1,268	0,76
Stimulation	0,746	1,67
Novelty	0,824	1,21

Table 7. Results of UEQ for Differ from Pilot 1 and 2

Co-funded by the Erasmus+ Programme of the European Union

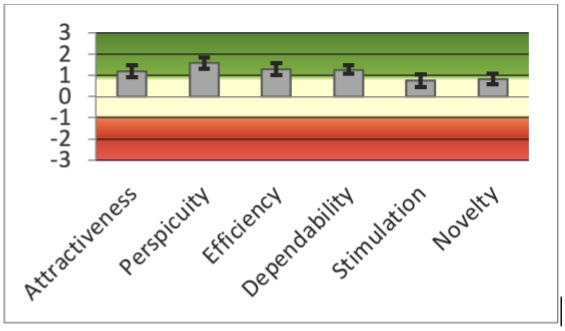
THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Pilot 1



Pragmatic and Hedonic Quality	
Attractiveness	1,02
Pragmatic Quality	1,43
Hedonic Quality	0,50

Pilot 2



Pragmatic and Hedonic Quality	
Attractiveness	↑ 1,19
Pragmatic Quality	↓ 1,38
Hedonic Quality	↑ 0,79

Figure 47. Results of UEQ for Differ from Pilot 1 and 2

3.3.4. Usability and User Experience of Hubert

In Pilot 2, evaluation of usability and user experience of the chatbot Hubert was also taken into consideration. Partners were searching for a short questionnaire that could be implemented at the end of the student's conversation with the chatbot, so that usability is measured during the onboarding process. We found a questionnaire UMUX-Lite as a good match for our goal since a preliminary research (Lewis et al, 2015) showed a close correspondence between UMUX-Lite and SUS scores.

UMUX-LITE (Lewis et al., 2013) is a two-item questionnaire based on the Usability Metric for User Experience (UMUX). One item is measuring perceived ease-of-use (Ease: "This system is easy to use") and the other is measuring perceived usefulness (Usefulness: "This system's capabilities meet my requirements") with seven-point scale items.

These two questions were reworded to correspond to the conversation style of the Hubert, and implemented into a Hubert educational template, after the question about satisfaction with Hubert:

Satisfaction with Hubert	How much did you enjoy using me (Hubert), from 1 to 5?
Adapted UMUX-Lite	Do my capabilities meet your requirements? Am I easy to use?

Ratings on UMUX-Lite scale was from 1 = Strongly Agree to 7 = Strongly Disagree.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

In pilot 2, 93 students had responded to the two items related to usability, 3 from Cyprus universities and 90 from the University of Zagreb (85 from the Faculty of Organization and Informatics, and 5 from the Faculty of Electrical Engineering and Computing). Data provided indicated that some students might have been confused with the reverse scale, in which 1 means the Strongly Agree or the highest score, while students are accustomed with 1 meaning the lowest score. Six responses were identified as outliers, so calculation was based on 87 responses. When calculating scores, the scale used in the conversation was reversed to provide correct calculations (1 - Strongly Disagree to 7 - Strongly Agree).

Based on those responses, perceived usability of Hubert was **71.1** on the scale from 0-100, which is average or good score. However, ease of use was rated much higher in contrast to the capabilities of Hubert. When calculating only one item, ease of use mean was 6.3 (st.dev 1.6) while capabilities mean was 4.2 (st.dev 2.2).

3.3.5. Results from the interviews with the students and educators

Analysis of the students' responses

In the scope of Pilot round 2, interviews with students were conducted. During courses, educators asked their students to participate in the interviews and the information about the interviews was shared on the portal of FOI's Student SUpport and Career Development Centre but nonetheless the number of respondents was lower than expected. Only **two students** have participated in the interviews.

Both students that have participated in the interviews are on the undergraduate level, one is attending Information Systems and Business Systems and the other Economics of Entrepreneurship study programme. Before Differ and Bo, both interviewees have encountered chatbots in their lives, mostly in the field of customer service, and in the interview they described their previous experiences and their experience with Bo.

Focus of the interview were: a) experience with chatbots and b) expectations related to chatbot usage. In the first part of the interview interviewers asked different questions related to positive and negative experiences with chatbots. Both interviewees **noted that their experiences were positive**, but they noted that they weren't completely satisfied with Bo's type of communication: one interviewee **didn't like the amount of text BO presented**, and the other **didn't like predefined answers** Bo was offering. Both interviewees would like to continue to use chatbots in their lives in the future. Both interviewees noted that a **bot's personality is a very important aspect** and that chatbots should have an easygoing and pleasant personality.

In the second part of the interview, the first interviewee is very positive about future chatbot usage: she thinks chatbots could be employed as customer service agents in education so that they could answer basic questions and she also notes that other fields (such as public services) could benefit

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

from chatbot implementation: this would help administrative staff to focus on other problems instead of answering the same questions over and over. She also thinks **it would be helpful if the chatbot could help her with learning course material**. Interviewee thinks that chatbots could help maintain online informal student communities: in her case students that used Differ created smaller communities where they discussed course material and talked about informal topics.

The second interviewee is a little bit sceptical about chatbot usage. Even though he would like to have an educational chatbot that would quiz him about the course material, remind him about his obligations such as exams, and facilitate team work, he **doesn't think that chatbots could help with maintaining online informal student communities especially since students have already connected through social media** and that they didn't perceive Differ as relevant.

SWOT analysis of educators responses

To get a detailed view on the functionalities and experience of usage of Differ and chat bot BO in various educational case scenarios , SWOT analysis was performed from responses of interviews with educators who participated in Pilot 2 (partners from Granada and RISE). SWOT Analysis is a useful framework for identifying **strengths, weaknesses, opportunities, and threats** of a chatbot application Differ. From the outcomes of this evaluation, we could identify main advantages and opportunities for other use case scenarios that could be reached in future in higher education. On the other hand, recognized weaknesses require better developed functionalities that will insure technical stability on all available operational systems and devices (desktop, mobile, browser version). Important challenges are also listed and should be tackled down to minimize risks that come from external environments including other alternative technical and communication solutions that are commonly used in higher education.

From the interviews, valuable feedback was collected in the mean of technical, usability and user experience improvements that will be taken into consideration before releasing new versions and functionalities of Differ and chatbot Bo.

Strengths	**Weaknesses**
• Availability in the official app stores of the mobile phones next to the official website, makes it very easy and simple for the user to install differ-chat. • The interface itself is very clear and very intuitive. • Bot "BO" is the first thing that pops up when you open the app, it is very helpful to understand how the app works and to ask what you don't understand.	• Differ requires installing a separate application / creating another account for the university. • Rejection to use the app caused by being told to download an application that will help students talk to strangers. • BO is present every time a user opens an application - it is not necessary and annoys users. • In the discover section there is no filter for hepping which new students you

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

• BO functions as an icebreaker between strangers are very useful. • BO as an initial guide to the app is really helpful. • The topics and suggestions that it makes to start the conversations are very well thought out, and it doesn't make us students fall into the typical patterns that we usually do. • The ability to create groups and subgroups of students can be useful for creating specific communities. The ability to insert gifs and the like in the chat has made conversations more fun and entertaining. • Quite a useful app, especially during the pandemic, when online classes make it difficult to meet people, especially in your first year at school. • There are two groups: the general one to receive information (channel) and the chat. • The ability to control the notifications of a topic in a community. • App offers good organization of media of every group in a separate "Files & Links" menu. • App offers the possibility of knowing new people who are close to your friends/university. • Safe place because the people in your community are real people who you may know or are familiar with. • An original way to meet people. Due to that reason, this bot will break the ice and also it is going to suggest new topics to speak about. • It is a good tool for synchronous communication. • Easy to send group and private messages.	want to meet or discover. • Technical limitations of application (not supported for Linux, landscape mode on iPad is not working). • Computer version of differ.chat does not have the same options as the mobile version. • Notifications don't work properly (at least on Xiaomi Smartphones). • Lack of usability and positive UX (e.g. clicking on intuitive areas for other platforms does nothing in differ.chat. For example, clicking on the name of a user in a chat sometimes does nothing. It doesn't show the information of what you're clicking or small profile photos which are not possible to maximize). • Shy people may feel obliged to start a conversation with the person that the bot has chosen for you. • Some users have fake names and profile photos and that may affect trust on the platform. • BO speaks English only which discourages people from non-english-speaking countries from continuing conversation. • The app should be translated in other languages • Hard to distinguish from other similar apps. • No added value when compared to the similar well established apps on the market. • The app is lacking features which are present in similar solutions on the market, such as: video calls, scheduled meetings, etc.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Opportunities	Threats
• Useful feature of the app would be the possibility of referencing messages from other messages, just as it can be done in WhatsApp or Telegram. • Aplication's users should primarily be students - tool used for fun enjoyment, not educators in academic context. • For successful use in academic context - app should be seriously related to work/scholar activities like existing apps integrated in LMS (e.g. Telegram). • Integration of the app of the first study year. • App can be improved and it could be an essential tool for Erasmus students. If you want to study abroad, like Erasmus, it would be interesting to meet new people from different universities before making a decision to choose one place or another and maybe to learn/improve foreign language with BO.	• Lack of interested users who will start and continue to use the app. • Promotion techniques don't work - new users are not attracted. • Many similar communication applications, even though they don't have chatbots, have long term users and good functionalities. • Students on many occasions found this application useless because they did not consider that BO would help them at all when they were told its advantages. • Existing learning tools and communication channels that universities use for years. • All information presented through Q&A is also available in official documents (ie. syllabus) and students do not wait for the access to the Q&A. • Students know themselves from previous study years, so there is no need for an introduction from a third party (ie. bot).

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

4. Pilot 3 courses

4.1. Best Practices in using chatbots – selected course case studies

4.1.1. Courses at the Cyprus Universities

Onboarding

In pilot 3 CYENS invited educators across different local universities to participate in the pilot, for testing Differ and/or Hubert. Two different invitations to join the pilot were shared with CYENS members' professional networks via emails. The first invitation concerned the use of Differ, while the second one the use of Hubert. Within the first invitation, improvements and new features added in the Differ chat were particularly stressed within the invitation email, as well as an introduction to BO, the social chatbot, that was a new addition in the Differ environment. Educators who were interested in participating in the Differ-BO pilot were required to complete a short online form (i.e., Google form, https://forms.gle/HCmmJ4tqtoyjehn3A) and provide their contact details. A total of 5 educators initially expressed interest to participate in pilot 3 for testing Differ. Within the second invitation concerning the advertisement of Hubert, the two different scenarios (course evaluation, self-reflection) were explicitly introduced in brief. Educators who were interested in participating in the Hubert pilot were required to complete a short online form (i.e., Google form, https://forms.gle/PYeg5XZ42zKU56E5A) and provide their contact details. A total of 4 educators initially expressed interest to participate in pilot 3 for testing Differ. CYENS researchers then proceeded with communicating with the educators via phone or personal emails, for providing more information about the pilot, the potential use of the chatbots in the class, and the data collection processes. Specifically, for the use of Differ, the following open document was shared with all educators, so as to guide them during the application of different scenarios (ice breaking activity, FAQs, split students into groups). Link to the document: Edubots Pilot Phase 3 – Instructions for Differ. Likewise, for the use of Hubert, the following open document was shared with all educators, so as to guide them during the application of different scenarios (informal course evaluation, students' self-reflection). Link to the document: Edubots Pilot Phase 3 – Instructions for Hubert. From the total of educators who initially expressed interest to participate in pilot 3, 4 educators finally participated in the pilot, summing up to a total of 7 different courses and 11 different use cases (two educators used Differ/ Hubert in more than one courses and for more than one use cases).

Best case

The following use case was one of the most successful cases of our pilots. Hubert was used in this course for two reasons: for supporting students' self-reflection and for facilitating informal

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

course evaluation (with the use of different evaluation templates). The use case was successful in the sense that: (i) most of the students interacted with Hubert in a fun and easy way, (ii) the interaction of students with Hubert increased the possibilities for self-reflection, (iii) students' responses were useful pieces of information for the educator to get insights into students' views about the course, the teaching methods applied, their understanding in course related topics and their expectations from the course.

Course name: History and Philosophical aspects of the Natural Sciences
Number of students: less than 50
Course description: The course attempts a mapping of the field of history and philosophy of science. Starting with a general overview of the history of science and the so-called "scientific method", we focus on the history of evolutionary thought, the historical relationship between science and religion, while we also consider what separates science from the non-scientific fields. We study science as a culture and a product of invention. As far as the philosophy of science is concerned, we are concerned with the issues of scientific explanation, realism, and the nature of theories. The section also includes the philosophy of physics, medicine, and social sciences, as well as specific examples of theories from the field of physics and their historical and philosophical implications.
Description of students' engagement with Hubert: Hubert was used in this class. The educator used Hubert 5 times during the semester, specifically: (i) once at the beginning of the semester for students' self-reflection, (ii) twice during the middle of the semester for self-reflection and course evaluation, and (iii) twice at the end of the semester for self-reflection and course evaluation. The links to those evaluations are provided below. For the self-reflection use case, the first interaction took place at the beginning of the semester aiming to assess students' expectations of the course. The second interaction took place in the middle of the semester, aiming to ensure that students are motivated, are on track, and whether they have any questions on certain topics, while the third interaction took place at the end of the semester, for evaluating students' self-perceived accomplishment of their learning goals. For the informal course evaluation use case, all interactions aim to the course evaluation. Through short, interactive dialogues with a chatbot, the students had the opportunity to provide informal feedback to their educator on their course. This feedback could be used by the educator for improving the applied teaching practices and approaches. Two templates were used for this use case. The first template was used during the academic semester, the second template, at the end. The items were essentially the same, it was just the wording that slightly differed in the two templates. The educator created each one of the five evaluations separately in Hubert.ai using the templates provided for course evaluation and students' self-reflection. Then the educator shared the generated link with the students, via a Learning Management Platform (Moodle). Then, the students could interact with Hubert by clicking on that link. Hubert posed the predefined (by the chosen template) questions to students, with a degree of smart interactions in the sense that Hubert could react positively or negatively to students' responses. Then, the educator reviewed the evaluation results, with students' responses, on her dashboard, having access to the raw data

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

(i.e., the actual conversation of each student with Hubert), but also to a thematic analysis of students' responses.

Affordances and challenges: Hubert was used by the students for self-reflection and course evaluation. Main affordances of the Hubert use: (i) easy process for the educator to set up an evaluation; (ii) easy way to invite students in interacting with Hubert, by sharing a link; (iii) interactive and fun way for students to evaluate the course, in an anonymous way and through a short interaction with Hubert, (iv) an alternative way for students to self-reflect on their previous knowledge of the course topic (at the beginning of the semester), on what they have learned and what are the main challenges they still face in the course, during the semester and at the end of the semester, respectively, (v) possibility for the educator to see and review raw data and easy access to them. In fact, reviewing the raw data was useful for the educator to identify the preferences and dislikes of students in relation to the delivery of the course, but also to get insights into students' misunderstandings or gaps in relation to the course content. Challenges faced in this course: (i) the fact that the Hubert-student interaction was taking place in the English language, hampered some students to initiate the conversation with Hubert, as they were not confident with the language, (ii) the educator was not completely satisfied with the thematic analysis of students' responses provided by the Hubert Evaluation Dashboard. For instance, some responses were clustered as having a negative connotation and were included in the negative aspects of the course evaluation, were those responses occurred from students' negative answers to the question "Do you want to add something more?". However the raw data provided insightful information to the educator about students' views, as mentioned above.

4.1.2. Courses at the University of Granada

Onboarding

In Pilot 3, Differ was used in the School of Informatics of Univ. Granada in 3 courses. Also, 2 other courses used Telegram for the first time. Results from pilots 1 and 2 already showed that it was very hard to encourage students to use or join new tools for communication with their peers or their educators. One of the first insights from previous pilots was that large groups for communities such as all the students in a Faculty or a School of Engineering are not particularly useful. The Univ. of Granada already provides institutional LMS platforms for this.

Building on this experience, communities for this pilot were set up for smaller groups (less than 50 students) and with a very specific purpose. In this case, for different courses of the Master in Big Data and Computer Engineering and for a course in Telecommunication Engineering. Also, in order to enrol more students, we decided to start using Telegram. This is a very popular tool in the IT sphere that was already used as an informal tool for communication among students.

Finally, in our case Hubert was not used after the experience in previous pilots since UGR already has other institutional (and some of them mandatory) mechanisms to collect feedback from students such as the institutional LMS platforms and a specific office for this purpose.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Best case

Course name: High-Performance Architectures for Computer Vision
Number of students: Less than 50
Course description*:* The course is part of the curriculum for the Master in Big Data and Computer Engineering. Its objective is to learn programming acceleration platforms such as FPGAs and GPUs for their use in Computer Vision processing. All students are 23 or older and most of them have part-time jobs in companies in the IT sphere. Their background is mostly Computer Engineering or Telecommunications.
Description of students' engagement with Differ and/or Hubert: The instructor created a community for this course. Students were invited to join via a link. The aim was for the students to have synchronous communication with the educator and with the other students in the course. Differ was used in the course as an additional communication tool (since students use an LMS for the course materials, evaluations, etc). Additionally, the educator in this course elaborated information of interest about the course such as responses to FAQs about the course, the toolsuites to be used in the practical labs, the evaluation of the course, etc.

Affordances and challenges: Students found with Differ a tool for synchronous communication with their educator and with the other students. From the educator's perspective, students also had the possibility to find the answers to frequently asked questions within the Differ community, relieving him from replying to repetitive questions and letting him focus on the educational process. However, students were not as active as expected, especially regarding the communication with their peers.

Overall opinion of educators participating in Pilot: The experience from educators was positive about the use of a tool that could relieve them from some frequent tasks and let them focus on more pedagogically relevant ones. However, the main result was that students were somehow tired or not very interested in using new tools. Especially since students were already using some familiar tools with the same objective. Remote learning during COVID could be one of the reasons that could explain this issue. The educator's determination was to take advantage of this for the next pilot and try to apply these tools with the same purpose (in this case, Telegram).

4.1.3. Courses at the University of Leeds

In Pilot 3 we trialled first use of Differ and Hubert by educators teaching Computing courses, but students did not take up these systems, and the few that tried gave negative feedback, saying that Differ and Hubert were not as usable as the standard chat and feedback tools provided by Leeds University IT service. In Pilot 2 we tried piloting with very different cohorts of students: instead of use of Differ in taught class chat, we offered Differ to Computing PhD students, and the mature students in the Lifelong Learning Centre. Computing PhD students who tried Differ were not impressed and quickly gave up, in preference to other platforms. The Lifelong Learning Centre students did take up Differ in larger numbers, but made very little use of chatbot. Neither cohort agreed to give feedback via Hubert, as Leeds has other existing standard feedback systems in place.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

For Pilot 3, we tried to revisit the use of Differ and Hubert with taught module classes, but with larger student numbers. We offered Differ and Hubert to students on 3 courses: Computing MSc module COMP5840 Data Mining and Text Analytics, class size 237; Computing undergraduate module COMP2121 Data Mining, class size 172; and Computing undergraduate module XJCO2121 Data Mining, taught remotely to 74 students at the Leeds-SWJTU Joint School campus in Chengdu, China. The educators were Eric Atwell, Artificial Intelligence professor and leader of the Leeds Edubots project team, and Yan Zhu, Artificial Intelligence professor and joint lecturer of XJCO2121 in Chengdu.

Onboarding

As the educators already knew about the EDUBOTS project, they did not require onboarding training on use of chat tools Differ and Hubert. Differ and Hubeet were offered by including them in the module Learning Resources in the Blackboard/Minerva virtual learning environment used for teaching at University of Leeds and the Leeds-SWJTU Joint School in Chengdu. However, Leeds University IT Service requires students and staff to use other platforms in online communication and issues its own course evaluation surveys. Hence we were unable to make the use of Differ or Hubert mandatory as part of a taught module class, or to include Differ and Hubert use in coursework for credit; and it was challenging to persuade students to volunteer to trial the use of optional platforms. A handful of students agreed to try Differ and Hubert, but they preferred the standard Leeds IT platforms for chat discussion and student feedback.

Best case

Course name: COMP5840 Data Mining and Text Analytics
Number of students: 237
Course description: 15-credit module taught over 11 weeks to computing students in MSc Advanced Computer Science and MSc Data Science and Analytics postgraduate programmes.
Description of students' engagement with Differ and/or Hubert: The course included natural language processing and text analytics applications, with chatbots cited as an example use case; students were invited to try Differ and Hubert as additional background study.
Affordances and challenges: Leeds University IT Service requires students and staff to use other platforms in online communication and issues its own course evaluation surveys. Hence we were unable to make the use of Differ or Hubert mandatory as part of a taught module class, or to include Differ and Hubert use in coursework for credit; and it was challenging to persuade students to volunteer to trial the use of optional platforms. A handful of students agreed to try Differ and Hubert, but they preferred the standard Leeds IT platforms for chat discussion and student feedback.
Overall opinion of educators participating in Pilot 3: We had not anticipated that Differ and Hubert would be competing with existing mandatory systems at Leeds University for student interaction and student feedback. We have advocated to senior management to trial alternative IT solutions for these IT functions, and Leeds University IT service has agreed to consider some

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

alternative systems for student-educator interaction and for eliciting student feedback; but not necessarily Differ and/or Hubert.

4.1.4. Courses at the University of Zagreb

Onboarding

In the third Pilot at the Faculty of Organization and Informatics (UniZg) Differ was used on 5 courses. Hubert was used in 2 courses but there was low interest due to various online faculty online surveys that were conducted in parallel. As most of the educators had previous experience with Differ, additional one-on-one instructions were provided to other educators who used Differ for the first time. The previous positive experience of the educator with the use of Differ is the main determinant for its reuse on the subjects.

Best case

For the best case course *Multimedia and web in business* was chosen. The course is elective in the summer semester of the 1st year of the Master's Degree at the university graduate study Economics of Entrepreneurship and had less than 50 students. Students were mostly full-time, interested in the course topics since the course is elective, but without much technological background. This course was chosen as the best case because of the educator's previous positive experience with Differ and tried out different strategies of employing Differ in the courses. In the courses in which she was an active Differ participant, sending messages, asking questions, the students were more active in using the Differ community. In the course where the educator was marginally active, there wasn't any activity from the students.

At the course Multimedia and web in business, students got the guidelines uploaded on the learning management system on how to join the **Differ community** and use the app. Over 90% of the students onboard the Differ. An educator created in Differ various topics for discussions with the students, related to the course structure (lectures, exercises, open chat). Also, Differ was used to further explain course announcements in LMS if needed. Students were actively participating in the discussions, mostly to get the answer to some exercise related problem or administrative questions (e.g will the lecture be recorded and available later in LMS, and similar). **Huber**t was used once in the last week of the semester to informally evaluate the course, but Response rate was low (only 7 responses or conversations were received).

Overall opinion of educators participating in Pilot 3

The previous positive experience of the educator with the use of Differ is the main determinant for its reuse on the courses. Students' lack of interest in new applications like Differ and sticking to existing communication solutions was the main challenge as the COVID-19 pandemic was active and online teaching and learning were in focus. In this situation the use of new tools was overwhelming both for students and educators.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

4.2. Interactions in Differ and Hubert

4.2.1. Analysis of system data in Differ

The Differ system data collected during Pilot 3 and the community categorisation were as described in section 2.4.1 Analysis of system data in Differ above, with additional events relating to users' interactions with the BO chatbot also captured.

The number of communities created and volume of users ever active in Pilot 3 are shown below.

Community Type	Pilot 3 Communities	Users Ever Active by Differ Community		
		Minimum	Maximum	Mean
Course	10	1	38	16.8
Programme	0	-	-	-
Department	0	-	-	-
Multi-Disciplinary	1	1	1	1.0
Total	**11**	**1**	**38**	**15.4**

Table 8.Number of communities and active users in pilot 3

4.2.2. Analysis of data from the course evaluations in Hubert

See section 5.2.2. Analysis of data from the course evaluations in Hubert below for analysis of Hubert activity across all pilot phases.

The "meta-evaluation" questions included to gather feedback on Hubert itself were expanded in several of the templates used from Pilot 3 onwards. However, some of these were presented with unintuitive Likert scale options for answers. For instance, Hubert asked some users "Am I easy to use? 1 = Strongly Agree, … 7 = Strongly Disagree".

These scales were adjusted in the templates used subsequently, but responses gathered with the initial wording are likely to be misleading due to the confusion caused.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

4.3. Results of the surveys and interviews about the chatbots

4.3.1. Differ and learning process

In the Pilot 3, an online survey was conducted to explore the role of chat application Differ and chatbot Bo in the learning process, and the perception of students about User Experience and Usability. All collected data was anonymous and voluntary. Data presented in this section represents summarised research findings from University of Zagreb, Faculty of Organization and Informatics, University of Granada and CYENS (former RISE) **who agreed on using the same questionnaire.** The University of Leeds did not participate in this survey. Research was conducted in the summer semester of the academic year 2020./2021.

In total, **116 responses** were collected from the students among three partners. After cleaning the data and excluding student's responses who did not use Differ in courses, **the sample consisted of 102 participants, N=102.** By home country, the participant's structure is presented on Figure 48.

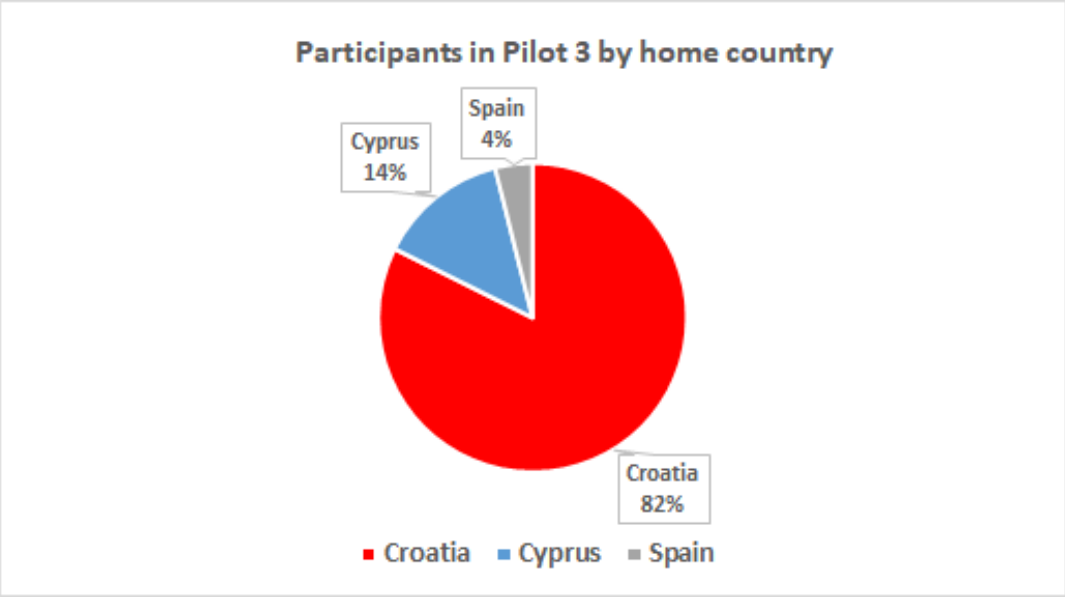


Figure 48. Participants by home country in Pilot 3, N=102

If we analyze demographic data about students, Figure 49 shows most of the students were male. By age, most of the respondents were 20-21 years old. Only 6,8% of students were older than 26 years.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

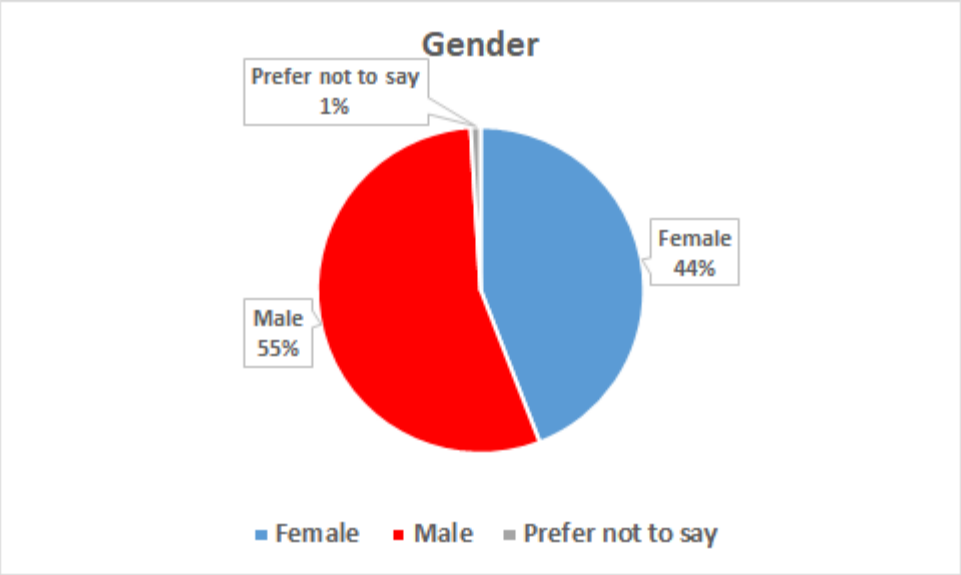


Figure 49. Gender distribution in Pilot 3, N=102

By the frequency of using Differ, most of the students reported that they used Differ several times a week (32,35%). 24,51% used it once a month during the semester.

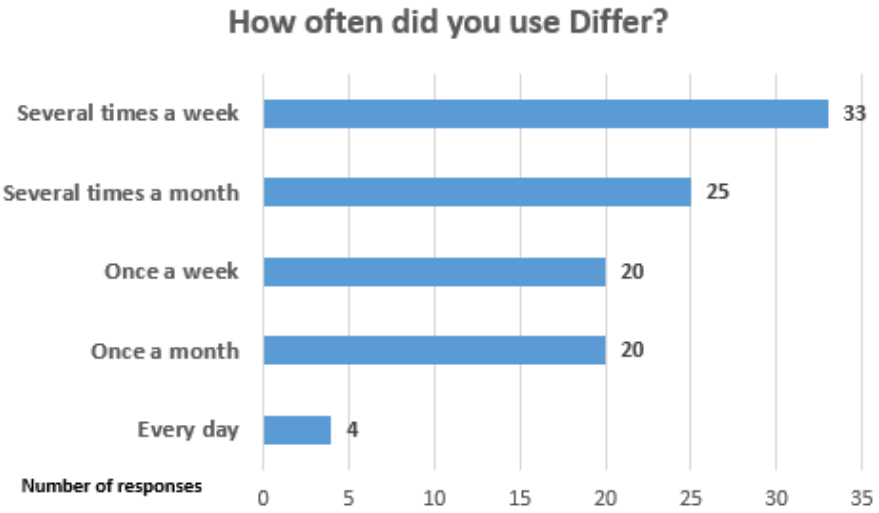


Figure 50. Frequency of Differ use in Pilot 3, N=102

From the Figure 51, it is obvious that most of the students used Differ via Mobile application (56,86%), but also great percentage of them , 37,35% used it as desktop application.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

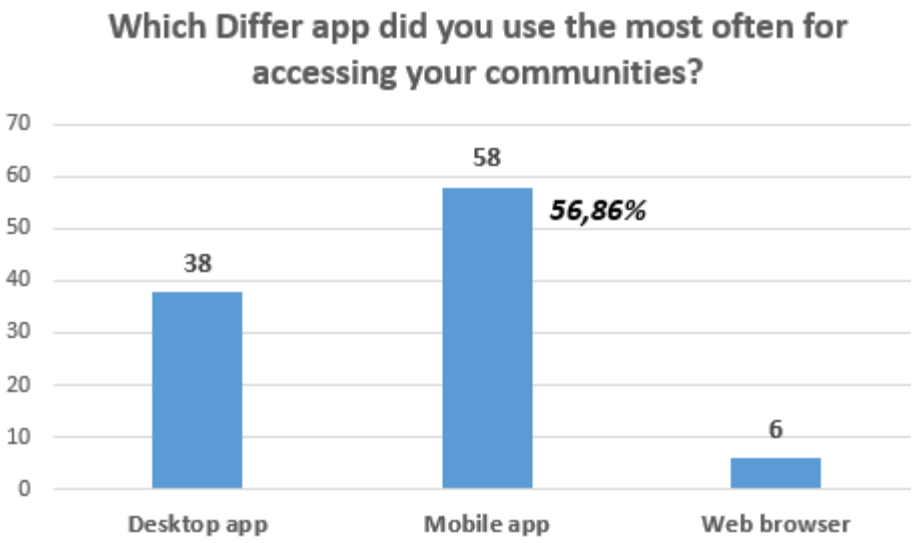


Figure 51. Access to Differ communities in Pilot 3, N=102

If we analyse student's answers related to the **Differ's role in learning process** Figure 52**,** following could be observed from the students answers:

- **40% of students** who filled out the survey felt that Differ **met their need as learners**, so we can consider it as an appropriate tool for enriching the whole process.
- **41% of student** reported that Differ helped them in **sharing knowledge with other students**, so we can consider it as appropriate tool for peer-to-peer collaboration and learning
- Most of the student**s have neutral** rather than positive or negative perceptions about means the Differ helped them in **successful completion of the course.**
- **Only ⅓** of respondents reported that Differ **helped them in improving their grade** on the course. Approximately 66% of them have a neutral attitude or disagree with this statement.
- Students reported more neutral and negative perceptions weather Differ helped them to improve their experience in the course

Analysis of students responses regarding **Differ as communication tool**, students reported more positive attitudes comparing to learning-supporting activities on the course:

- For **68,75%** students Differ helped them to have **better connection with course educators 53,13%** students **felt free** to share their feelings and expressions in Differ
- Aproximately 40% of students reported that Differ helped them in connecting with other students

In Pilot 3 form 3 partner countries , from the student answers Differ can be considered **as a safe communication platform** where students can express their feelings and opinions, ask questions about the course, but also have more informal yet constructive vhats with course educators and peers. Differ scored better as a communication tool rather than a tool that supports learning

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

activities on the course, but almost ⅓ students in the research reported that **Differ somehow supported their learning engagement** in various activities and collaboration with students during the COVID-19 marked semester.

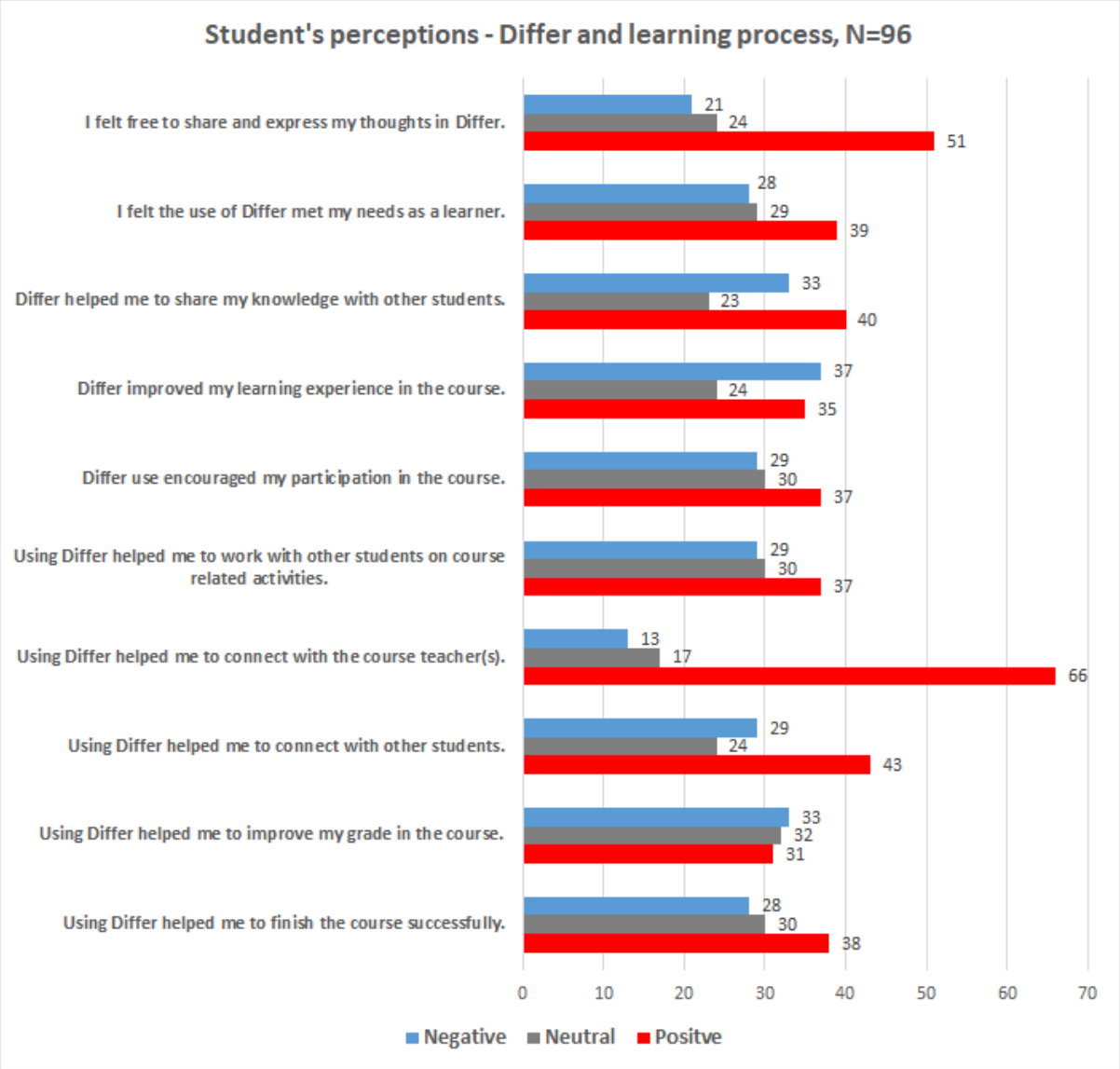


Figure 52. Distribution of students' opinions related to the 10 learning support statements, N=96

4.3.2. Differ and teaching process

Educators' Survey About Differ and Hubert Use

In the Project as the part of Pilot 3 phase we performed a Survey on usage of Differ, BO and Hubert on the courses with educators. The goal was to find out how educators use Differ, Hubert

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

and especially BO as well as how it helps them in the teaching process and connecting with students. **Only 9 educators** participated in the survey so all quantitative data will not be further interpreted, but just presented as is.

In total, 6 female and 3 male educators participated in the survey. 4 of them were from Croatia and two from Spain and Cyprus. Most of the respondents are associate professors, following 2 Associate Professors and two teaching assistants. Related to course level, 4 courses are on the Masters level of study and 5 on the undergraduate level. Students enrolled in courses were **dominantly full time students.** What is common to **6 subjects** is that they are **related to IT in a certain way** (e.g. data mining, multimedia, etc. and had some kind of practical exercises in course, while one subject was focused on developing students' communication skills. **Teaching activities** that were provided in the courses were lectures (9 responses), above mentioned practical exercises (7 responses) and seminars (6responses). 22,2 % of educators reported that 21-30% of enrolled students will have to take the final exam by the end of the semester, and **77,8% of them expect rates 0-10% of students not successfully completing the course,**

57,1% of educators used Differ only on **one course and 28,6%** of them on two courses. Most of them used Differ **several times a week to communicate with students** (Figure 53), usually via mo**bile application** (57,15%). The same percentage of educators were invited to use Differ by other colleges, and just one educator via invitation at the EDUBOTS Differ page.

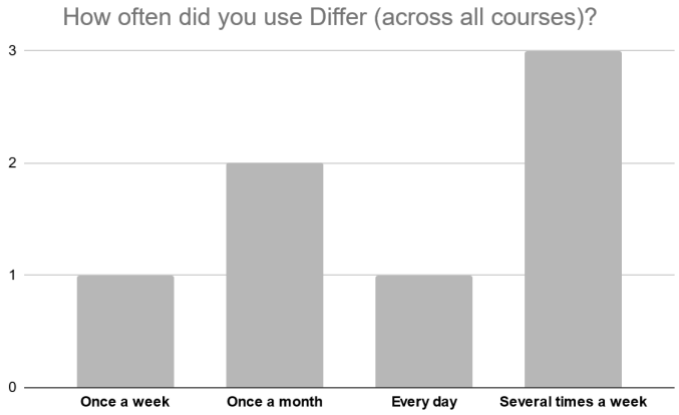


Figure 53. Frequency of Differ use

Regarding the impact of Differ on courses (Figure 54), **6 educators did not recognize that it had some impact and for 3 of them it had positive outcomes.** Some of the arguments for such an opinion are that Differ helps educators to connect with students during the COVID-19 pandemic, but the majority of students who onboard on Differ didn't use it regularly. Therefore, no impact could be accomplished. One educator reported that Differ helped in activities like sharing teaching materials.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

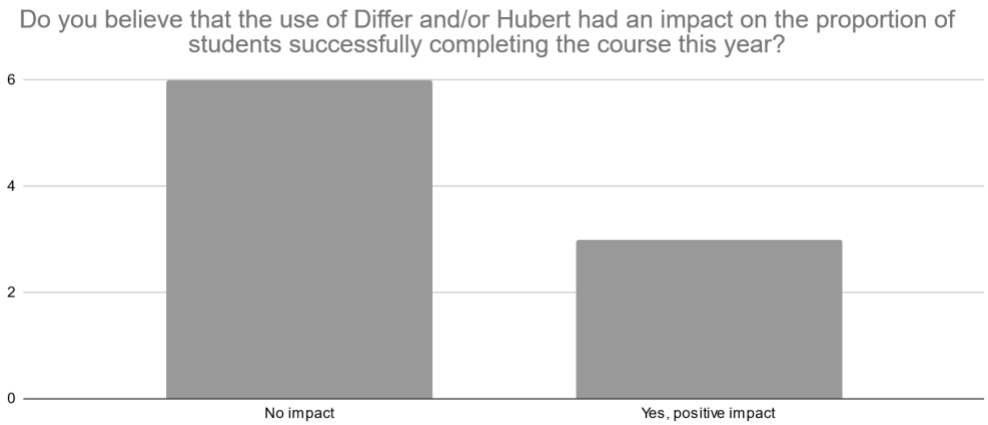


Figure 54. Impact of chatbot applications use

Majority of the educators **(86,6 %) used chatbot BO to create a course community** and just in one case others did it for them (Figure 55). If we analyse other options of chatbot BO, the most popular functionality was Frequently Asked Questions (FAQs) for 57,1% of educators. Distributions for outher otions are on Figure 56. None of the educators reported use of student mentors in Pilot 3.

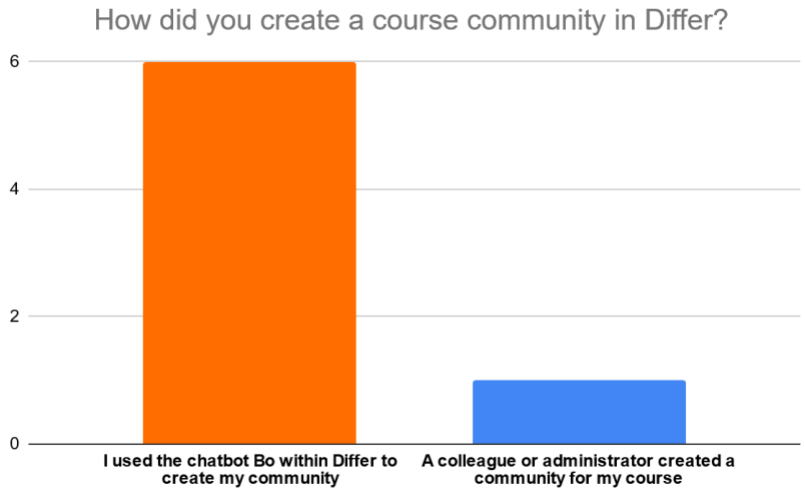


Figure 55. Creating course community in Differ

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

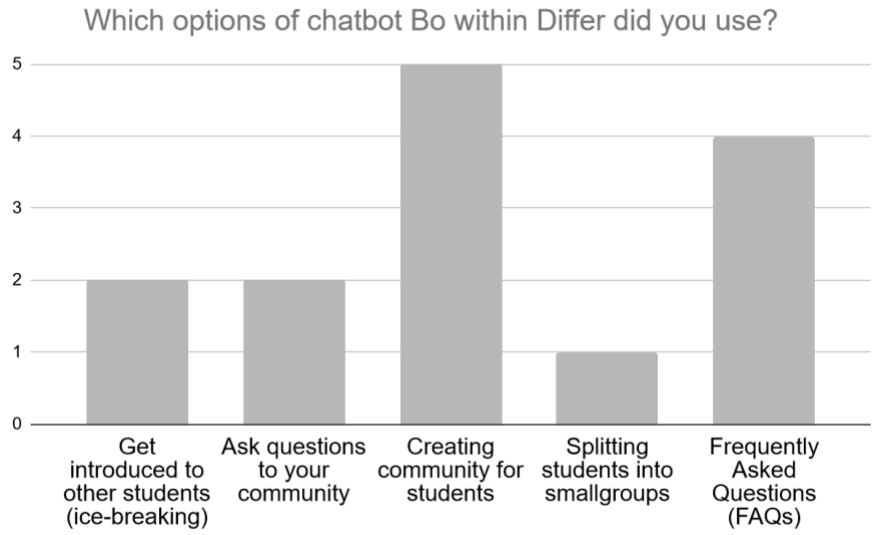


Figure 56. Which options of chatbot Bo within Differ did you use?

In Pilot 3, educators mostly used Differ to discuss various topics with students, course announcements, to prepare students for project activities and explaining course materials (Figure 57). It is obvious that educators were oriented **strongly on communication activities** and to stay connected with students in online teaching because of COVID-19. Educators reported that Differ reduced the number of emails and made communication faster and easier in the online teaching environment.

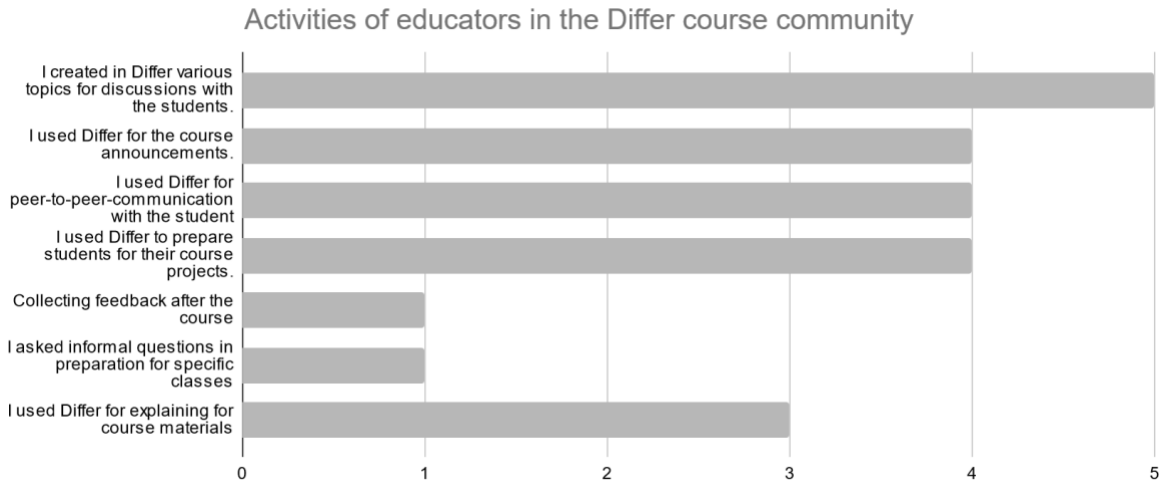


Figure 57. Educators' activities in Differ

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

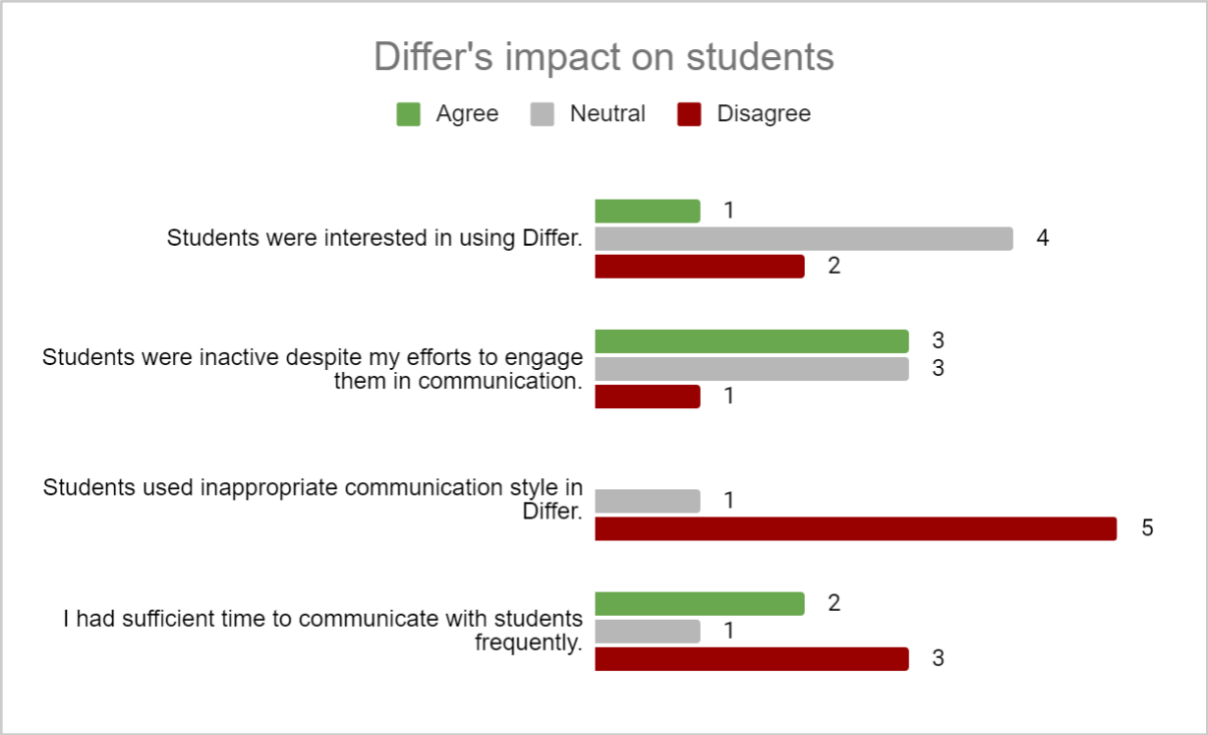


Figure 58. Differ's impact on students

In Pilot 3, educators mostly perceived students as inactive in the Differ, dhespite their efforts to motivate them. This is further confirmed through neutral perception towards the Differ app (Figure 58). One of the reasons why is that some educators did not have enough time to communicate with the students on a regular basis. On the other hand, students did not use inappropriate language in communication when they were using Differ, which is in line with the academic values.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

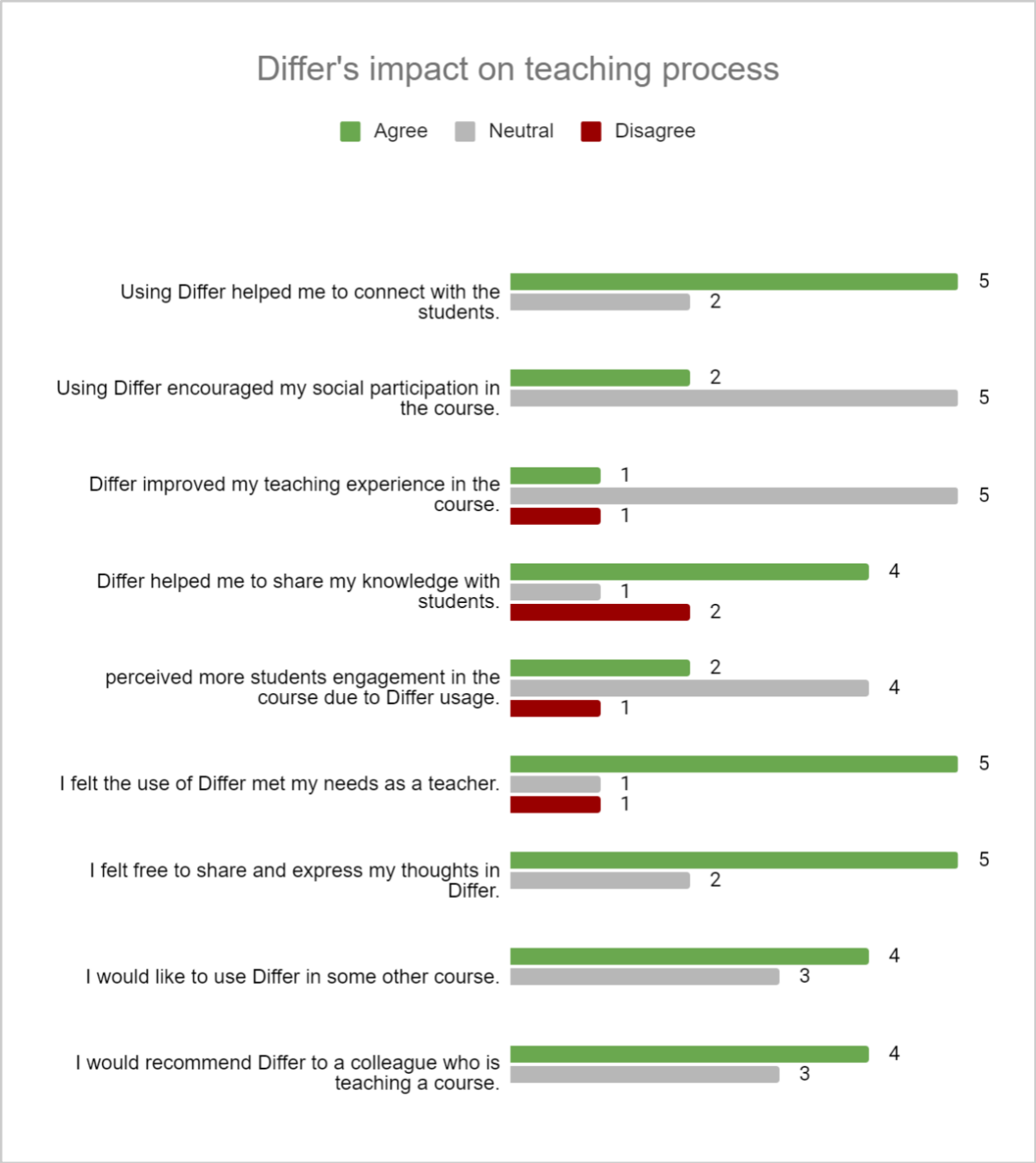


Figure 59. Differ's impact on teaching process

In Pilot 3 educators mostly perceived Differ's impact on the teaching process as positive or neutral (Figure 59). Educators' perception was that, despite lack of time, Differ helped them connect with students. On the other hand, educators had a neutral stance towards improvement of teaching experience, social participation in a course, which is in line with the perception that students were also rather inactive in the Differ. Based on the data presented in the figure, we can conclude that

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

educators were free to share their thoughts, would recommend using Differ to their colleagues, and would mostly like to use Differ in other courses.

Regarding the use of **chatbot Hubert for course evaluations** three educators used Hubert on a single course, while three used Hubert on two courses. Four of them used Huber once, one educator used it two times and one used it three times.

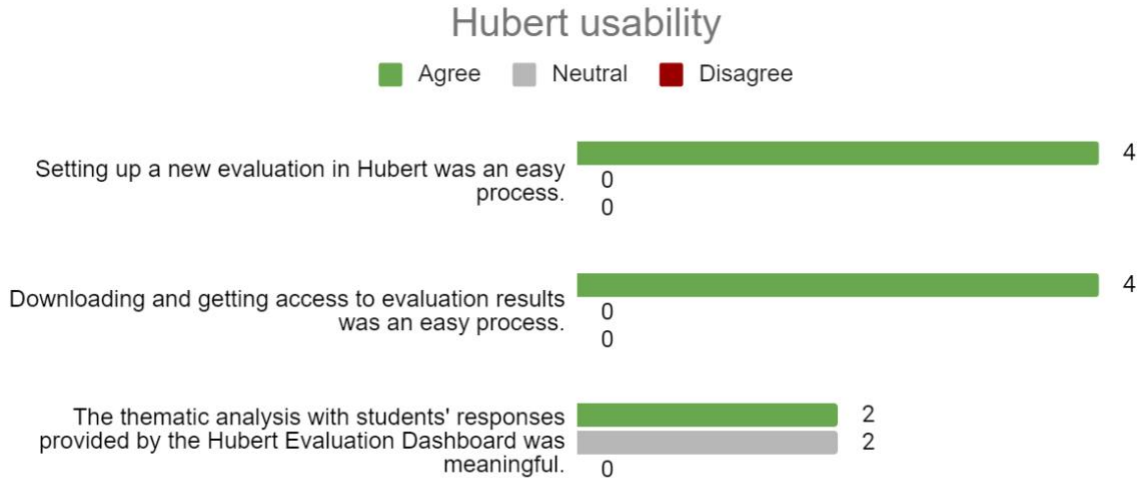


Figure 60. Hubert's usability

In Pilot 3 educators dominantly perceived Hubert positively (Figure 60). All agree that Hubert is easy to set up, as well as it is easy to access and download evaluation data. Only aspect which is evenly split between positive and neutral opinion is thematic analysis which is generated within the Hubert.

4.3.3. Usability and User Experience of BO within Differ

In Pilot 3, focus was put into evaluation of usability and user experience of chatbot BO within Differ, since BO was introduced as a new feature in Pilot 2 but hasn't been evaluated separately but only as an integral part of Differ. Several evaluations were performed in Pilot 3: usability of BO with SUS questionnaire (Bangor et al., 2008; Hassenzahl & Tractinsky, 2006), usability of BO evaluated with a questionnaire specifically created for chatbots: Chatbot Usability Questionnaire, CUQ (Holmes et al., 2019), and user experience of BO with User Experience Questionnaire, UEQ (Schrepp et al., 2017; User Experience Questionnaire (UEQ), n.d.) used in Pilot 2.

Usability of BO evaluated with SUS

Only 23 students responded to the SUS questionnaire in Pilot 3, 11 from the Cyprus universities and 12 from the Faculty of Organization and Informatics, so further results should be interpreted

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

with caution. To have a margin of error of +/- 5,00 points with 90% confidence level, we should have a sample size of 24.

Pilot 3 - BO	N=23
Mean SUS score	**62,5** ↓
StDev	0,78

Pilot 2 - Differ	N=71
Mean SUS score	**66,8** ↓
StDev	0,84

In comparison to usability of Differ in Pilot 1 (Mean SUS score 72,4) and Pilot 2 (Mean SUS score 66,8), usability of BO measured with SUS was perceived lower than Differ: 62,5. On the grade scale, this belongs to D grade and marginal acceptance of the users to use it further (see figure Figure 61). This result calls for serious redesign decisions which should be informed by additional user research.

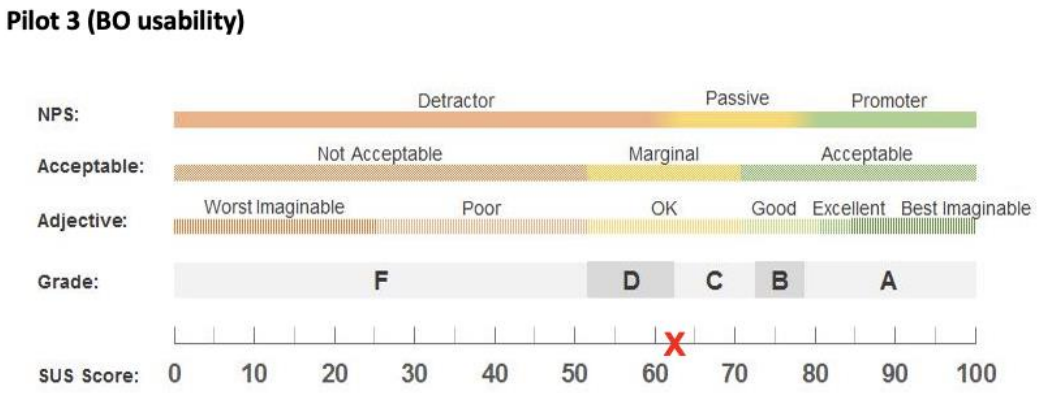


Figure 61. BO's usability on various scales (Pilot 3)

Usability of BO evaluated with CUQ

Additional questionnaire was used to further investigate the opinions about the specifics of BO. Chatbot Usability Questionnaire (CUQ) is based on the Chatbot Test tool (Martín et al., n.d.) which assesses seven categories of chatbot's design: personality, onboarding, navigation, understanding, answering, error management and intelligence of a chatbot. The CUQ is designed to be comparable to SUS with addition of statements related to chatbots, and includes 16 items. Participants' levels of agreement with sixteen statements are ranked out of five, from "Strongly Disagree", to "Strongly Agree" (Holmes et al 2019).

Altogether, a hundred participants have evaluated BO with CUQ: Cyens 14 from Cyprus universities, 2 from University of Granada, and 84 from the Faculty of Organization and Informatics. Their mean answers on the scale from 1 to 5 are presented in the Figure 62.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

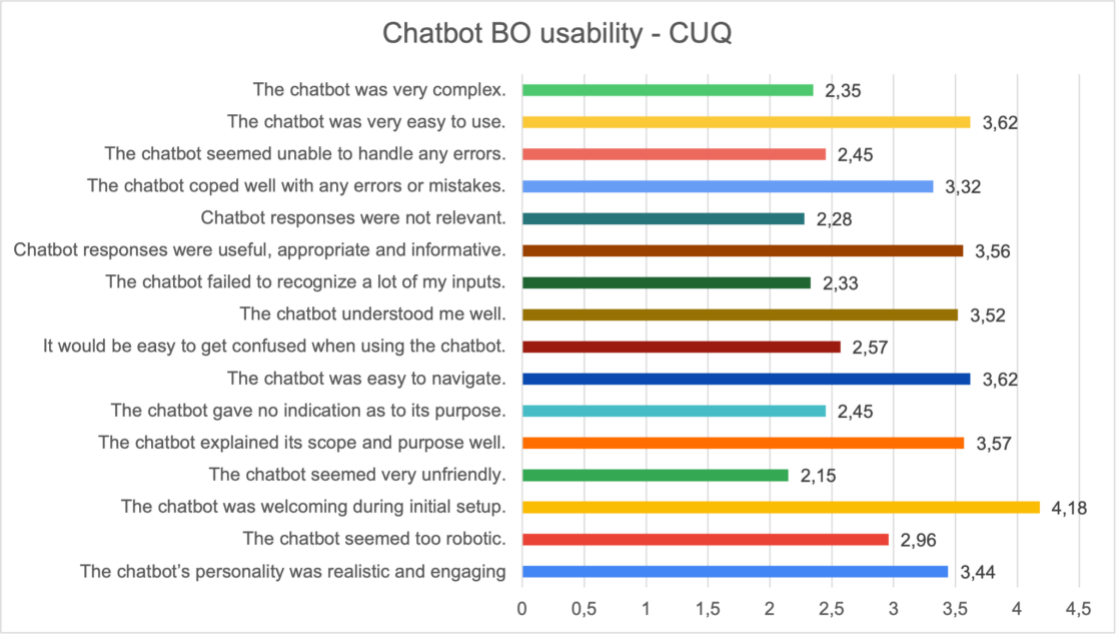


Figure 62. Usability of BO evaluated with CUQ (Pilot 3)

Scores **better than average** were achieved for: The chatbot was welcoming during initial setup (4,18), The chatbot was very easy to use (3,62) and Chatbot responses were useful, appropriate and informative (3,56). The lowest score was for the statement "The chatbot seemed very unfriendly", but since this is a negative statement, BO was actually perceived as a **friendly chatbot**.

User experience of BO evaluated with UEQ

In Pilot 3, BO was evaluated with 26 pairs of a positive or negative attribute on a differential scale within UEQ. Overall, **91** students (85 from Croatia, 6 from Cyrus) responded to the questions which evaluated *Attractiveness* (gives overall impression about BO), *Perspicuity* (measures learnability of using BO), *Efficiency* (measures efficiency of interaction with BO), *Dependability* (measures levels of control of BO), *Stimulation* (measures levels of motivation to use BO) and *Novelty* (measures levels of innovation or originality of BO).

Analysis showed that BO received a more or less **neutral evaluation** on the scale form -3 (horribly bad) and +3 (extremely good), on which values > 0,8 represent a positive evaluation. Only *Perspicuity* (learnability) was evaluated **positively** (mean 0,836) as seen on the Figure 63. Stimulation is the category that received the lowest score. Both pragmatic and hedonic quality, as well as attractiveness were evaluated neutral.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Pilot 3 - BO		N=91
UEQ Scales (Mean and Variance)		
Attractiveness	0,622	2,32
Perspicuity	0,836	1,82
Efficiency	0,577	2,17
Dependability	0,720	1,34
Stimulation	0,199	2,33
Novelty	0,613	1,86

Pilot 3 - BO	N=91
Pragmatic and Hedonic Quality	
Attractiveness	0,62
Pragmatic Quality	0,71
Hedonic Quality	0,41

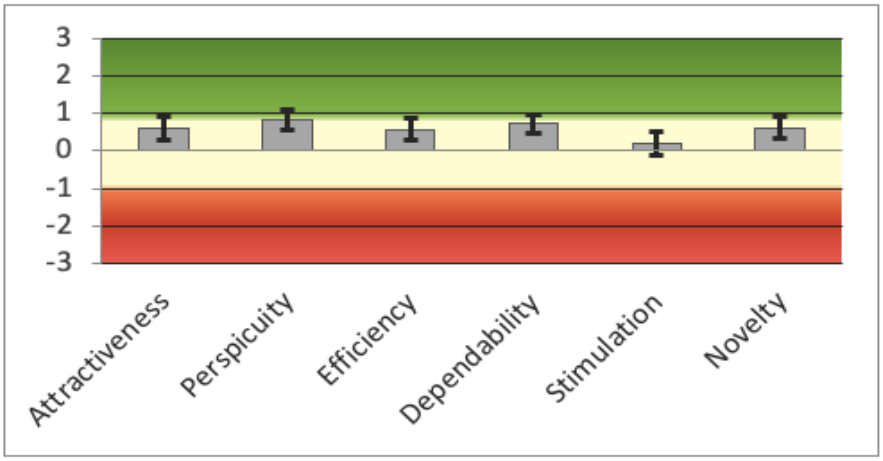


Figure 63. User experience of BO evaluated with UEQ (Pilot 3)

Benchmark revealed (see Figure 64) that BO was perceived **below average** or even bad application in comparison with data from 20190 persons from 452 studies concerning different products (business software, web pages, web shops, social networks). According to those results, both functionalities and user interface should be substantially improved to be attractive to users, provoke positive emotions and willingness to use the application. However, those results derived from respondents mainly from one faculty which might skewed the results into a negative direction. Unfortunately, students from other universities were not motivated to fill-in the surveys.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Scale	Mean	Comparisson to benchmark	Interpretation
Attractiveness	0,62	**Bad**	In the range of the 25% worst results
Perspicuity	0,84	**Below Average**	50% of results better, 25% of results worse
Efficiency	0,58	**Bad**	50% of results better, 25% of results worse
Dependability	0,72	**Bad**	In the range of the 25% worst results
Stimulation	0,20	**Bad**	In the range of the 25% worst results
Novelty	0,61	**Below Average**	50% of results better, 25% of results worse

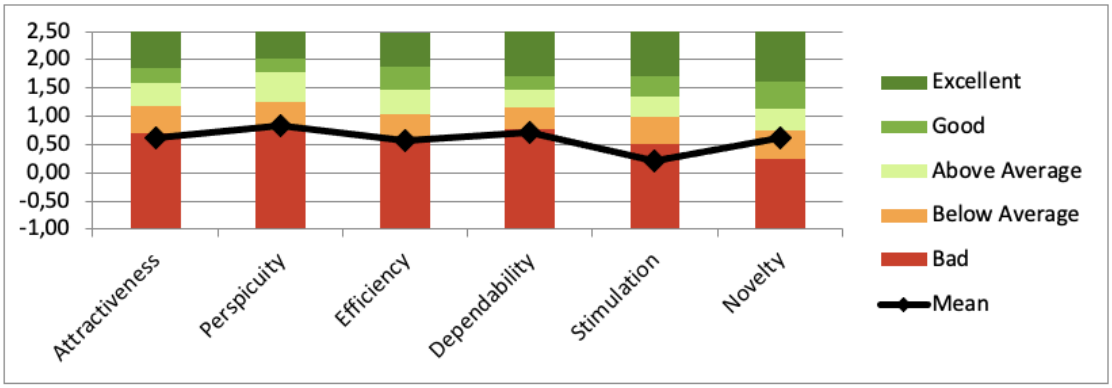


Figure 64. User experience of BO evaluated with UEQ (Pilot 3)

4.3.4. Usability and User Experience of Hubert

Since usability and user experience of Hubert was not evaluated so extensively in previous pilots as Differ (and BO), we decided to evaluate Hubert with additional questionnaires implemented in one Google Form: Chatbot Usability Questionnaire, CUQ (Holmes et al 2019) for evaluating chatbot usability and User Experience Questionnaire, UEQ (Schrepp et al., 2017; User Experience Questionnaire (UEQ), n.d.) for evaluating user experience. As in Pilot 2, short evaluation of usability of the chatbot Hubert was performed with the questions from the questionnaire UMUX-LITE (Lewis et al, 2013) integrated into conversation with Hubert.

Usability of Hubert evaluated with UMUX-LITE

With the question “Am I easy to use?” perceived ease-of-use was measured, while with the question “Do my capabilities meet your requirements?” perceived usefulness was measured, both with seven-point scale items. Those two questions were integrated at the end of the conversation with Hubert.

In Pilot 3, 73 students had responded to the two items related to usability, 51 from the University of Leeds and 22 from the University of Zagreb, Faculty of Organization and Informatics. Data provided still indicated that some students might have been confused with the reverse scale (1 - Strongly Agree to 7 - Strongly Dissagree). Four responses were identified as outliers, so calculation was based on 69 responses. When calculating scores, the scale used in the

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

conversation was reversed to provide correct calculations (1 - Strongly Disagree to 7 - Strongly Agree).

Based on those responses, perceived usability of Hubert in Pilot 3 was **71.9** on the scale from 0-100, which is average or good score, and only 0.8 point better than in Pilot 2 score.

However, perceived ease of use was rated much higher in contrast to the capabilities (perceived usefulness) of Hubert. Similar to the results from Pilot 2, ease of use mean was 6.1 (st.dev 1.5) while capabilities mean was little bit higher and scored 5 (st.dev 2.0).

Usability of Hubert evaluated with CUQ

To further investigate the opinions about the chatbot specifics of Hubert, a Chatbot Usability Questionnaire (CUQ) was used. The CUQ is designed to be comparable to SUS with addition of statements related to chatbots, and includes 16 items. Participants' levels of agreement with sixteen statements are ranked out of five, from "Strongly Disagree", to "Strongly Agree" (Holmes et al 2019).

Altogether, 42 participants have evaluated Hubert with CUQ: 17 from Cyprus universities, and 25 from the Faculty of Organization and Informatics. Their mean answers on the scale from 1 to 5 are presented in the Figure 65.

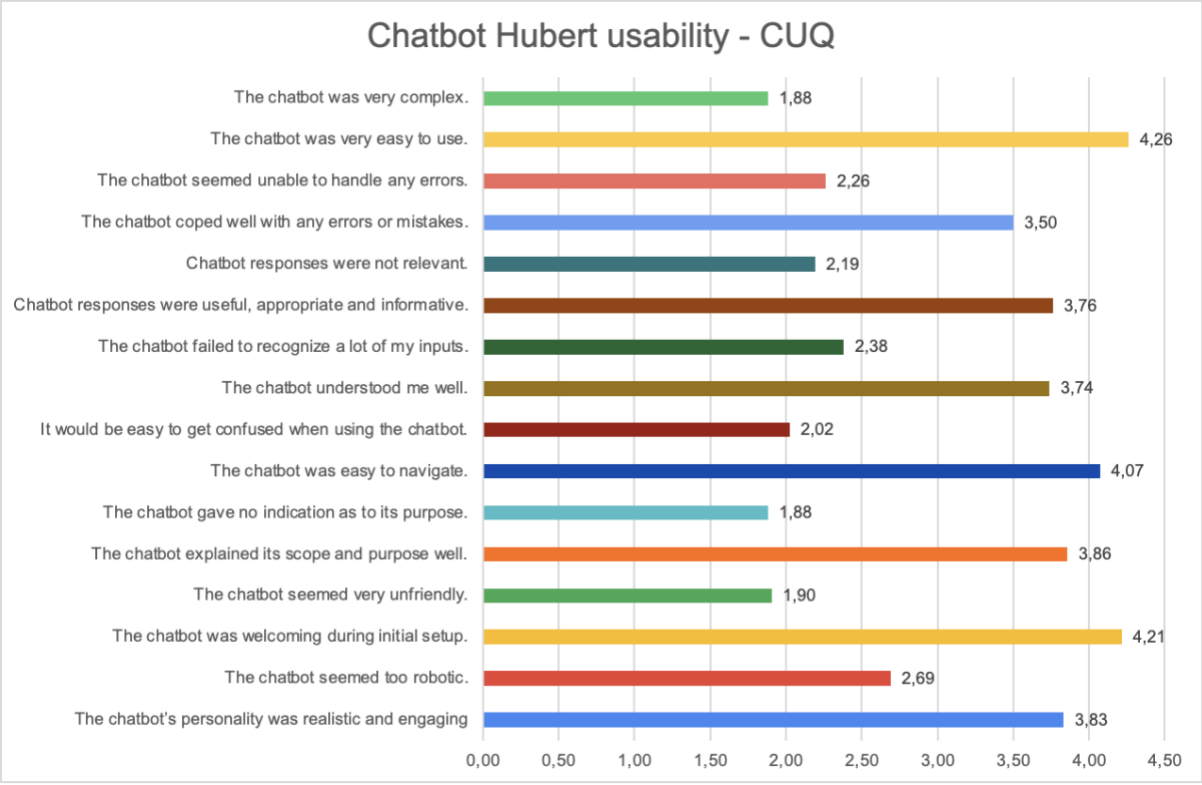


Figure 65. Usability of Hubert evaluated with CUQ (Pilot 3)

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Scores **higher than 4** (Agree with the statement) were achieved for: The chatbot was very easy to use (4,26), The chatbot was welcoming during initial setup (4,21), and The chatbot was easy to navigate (4,07). The lowest scores were received for the negative statements "The chatbot was very complex." and "The chatbot gave no indication as to its purpose.", so Hubert was actually perceived as a **simple chatbot with a defined purpose**.

User experience of Hubert evaluated with UEQ

In Pilot 3, Hubert was evaluated with 26 pairs of a positive or negative attribute on a differential scale within UEQ. Overall, 25 students from Croatia responded to the questions which evaluated *Attractiveness* (gives overall impression about Hubert), *Perspicuity* (measures learnability of using Hubert), *Efficiency* (measures efficiency of interaction with Hubert), *Dependability* (measures levels of control of Hubert), *Stimulation* (measures levels of motivation to use BO) and *Novelty* (measures levels of innovation or originality of Hubert). This response rate is rather low and for Precision=0.5, Err.Prob.=0.05, response rate should be above 30 (20 for *Dependability*), so the results should be interpreted with caution.

Analysis showed that Hubert received **positive evaluation** on the scale form -3 (horribly bad) and +3 (extremely good), on which values > 0,8 represent a positive evaluation. Only *Stimulation* (motivation) was evaluated **netral** (mean 0,690) as seen on the Figure 66. Both attractiveness and pragmatic quality were evaluated positively, while hedonic quality was evaluated neutral.

Pilot 3 - Hubert	N=25	
UEQ Scales (Mean and Variance)		
Attractiveness	1,127	2,38
Perspicuity	1,360	1,99
Efficiency	1,170	2,13
Dependability	1,160	1,28
Stimulation	0,690	2,44
Novelty	0,890	2,04

Pragmatic and Hedonic Quality	
Attractiveness	1,13
Pragmatic Quality	1,23
Hedonic Quality	0,79

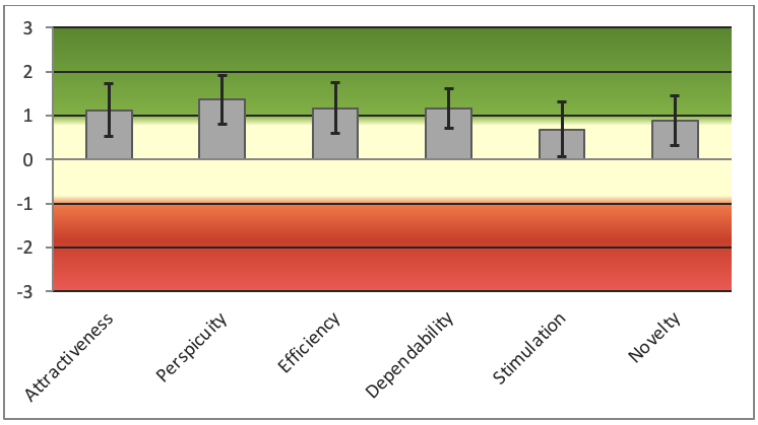


Figure 66. User experience of Hubert evaluated with UEQ (Pilot 3)

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Benchmark revealed (see Figure 67) that Hubert was perceived mostly **above average** or below average for *Attractiveness* and *Stimulation*, in comparison with data from 20190 persons from 452 studies concerning different products (business software, web pages, web shops, social networks). However, those results derived from only 25 respondents so results might be different if more students responded. Unfortunately, students from other universities were not motivated to fill-in the survey.

Scale	Mean	Comparisson to benchmark	Interpretation
Attractiveness	1,13	**Below average**	50% of results better, 25% of results worse
Perspicuity	1,36	**Above Average**	25% of results better, 50% of results worse
Efficiency	1,17	**Above Average**	25% of results better, 50% of results worse
Dependability	1,16	**Above Average**	25% of results better, 50% of results worse
Stimulation	0,69	**Below Average**	50% of results better, 25% of results worse
Novelty	0,89	**Above Average**	25% of results better, 50% of results worse

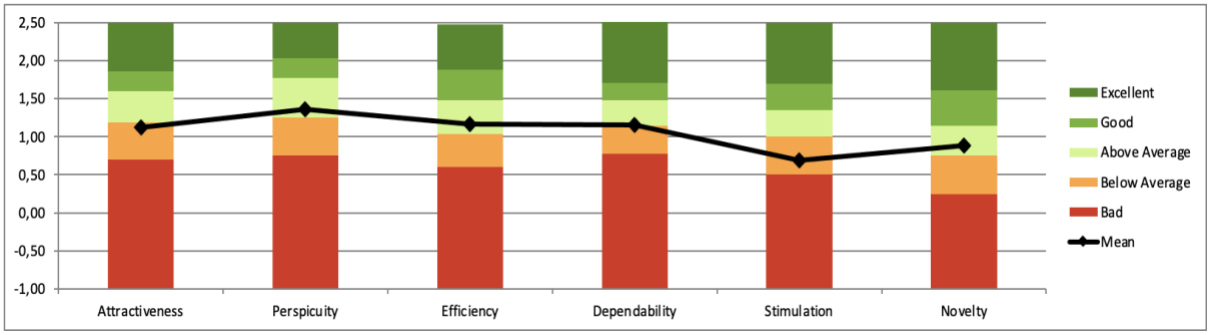


Figure 67. User experience of Hubert evaluated with UEQ (Pilot 3)

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

5. Pilot 4 courses

5.1. Best Practices in using chatbots – selected course case studies

5.1.1. Courses at the Cyprus Universities

Onboarding

In pilot 4, CYENS followed a similar strategy, as in the previous pilots. CYENS invited educators across different local universities to participate in the pilot, for testing Differ and/or Hubert. In this pilot, one single invitation was sent to CYENS members' professional networks via emails. Educators who were interested in participating in the pilot were required to complete a short online form (i.e., Google form, https://forms.gle/AcxFJaRnSDixsmHV8) and provide their contact details. A total of 5 educators initially expressed interest to participate in pilot 4 for testing Differ or Hubert, out of which 4 finally participated in the pilot. CYENS researchers then proceeded with communicating with the educators via phone or personal emails, for providing more information about the pilot, the potential use of the chatbots in the class, and the data collection processes. Emails with particular instructions on how to start using the two chatbots was also sent. In addition, educators were prompted this time to join the cMOOC community. Educators also received a set of manuals with guidelines for the use of Differ and Hubert. Specifically, for the use of Differ, the following open document was shared with all educators, so as to guide them during the application of different scenarios (ice breaking activity, FAQs, split students into groups). Link to the document: Edubots Pilot Phase 3 – Instructions for Differ. Likewise, for the use of Hubert, the following open document was shared with all educators, so as to guide them during the application of different scenarios (informal course evaluation, students' self-reflection). Link to the document: Edubots Pilot Phase 3 – Instructions for Hubert. Four educators finally participated in the pilot, summing up to a total of 6 different courses (one educator used Differ and Hubert in two courses and another educator used Hubert in two courses).

Best case

The following use case was also a successful case of our pilots. Hubert was used in this course for supporting students' self-reflection. The use case was successful in the sense that: (i) all the students interacted with Hubert in a fun and easy way, (ii) the interaction of students with Hubert increased the possibilities for self-reflection, (iii) students' responses were useful pieces of information for the educator to get insights into students' views about their expectations from the course and their understanding in course related topics.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Course name: Current Trends in the Teaching of Mathematics and Science

Number of students: Less than 50

Course description: The aim of the course is to bring students in a first contact with modern research in Mathematics Education and Science Education. The course examines selected areas of research in both areas. Students to develop: Basic thinking and scientific method skills, personal philosophies for learning and teaching mathematics and science, basic research skills related to the teaching of mathematics and science, to develop knowledge for modern learning theories and didactic approaches in Natural Sciences and Mathematics, to develop argumentation skills, to be informed about didactic approaches related to the nature of science and mathematics, to plan didactic activities through the prism of research findings in cognitive psychology and didactics of Science and Mathematics, to gain knowledge and develop skills which will be the basis for the creation of personal philosophies of teaching and learning for Science and Mathematics.

Description of students' engagement with Differ and/or Hubert: The educator used Hubert in the middle of the academic semester as part of a series of activities dedicated to assessment in science teaching. As part of this activities, Hubert was introduced to the students as an example of technological tools that can be used in the class with students for self-reflection on their learning progress and for an active engagement of students. The educator in this case created a new evaluation in Hubert.ai using the template provided for students' self-reflection. Then the educator shared the generated link with the students via the Moodle platform and as part of students' weekly assignments. All the students in the class interacted with Hubert by clicking on that link. Hubert posed the predefined (by the chosen template) questions to students, with a degree of smart interactions in the sense that Hubert could react positively or negatively to students' responses. Then, the educator reviewed the evaluation results, with students' responses, on her dashboard, having access to the raw data (i.e., the actual conversation of each student with Hubert), but also to a thematic analysis of students' responses. Concluding, Hubert was used in this course for self-reflection and as an example of technological tools that can be used by the students themselves in their own classes.

Affordances and challenges: Since the activity was obligatory for the students, as part of their weekly assignments, all students in the cohort interacted with Hubert, thus providing a good pool of responses for the educator. Also, it was a good opportunity for the students to experience a technological tool that can be used in their own classes with their own students. Main affordances reported in the previous pilots, such as (i) easy process for the educator to set up an evaluation; (ii) easy way to invite students in interacting with Hubert, by sharing a link are still applicable in this case as well. There was only one challenge, which was also reported previously: the fact that the educator was not completely satisfied with the thematic analysis of students' responses provided by the Hubert Evaluation Dashboard.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

5.1.2. Courses at the University of Granada

Onboarding

In pilot 4, UGR used a hybrid approach developing chat groups in Differ and as done in pilot 3, also in Telegram in order to encourage students using the chatbots, given the poor results in previous pilots in terms of enrollment. So in this pilot, 1 course was piloted with Differ and the other 3 with Telegram. UGR still invited educators to use Differ at our university and specially for the School of Informatics (mainly for the Computer Science and Telecommunication Engineering educators, but also educators for the masters in Computer Science and Technology and the education master).

Best case

Course name: Electronic Digital Systems

Number of students: 50 to 100
Course description: This course is part of the curriculum for Telecommunication Engineers (third year). The main objective is to teach the hardware programming language VHDL for reprogrammable hardware platforms (Field programmable gate arrays or FPGAs). Students take an exam at the end of the semester for the theoretical contents and another one for practical experiments. This is 50% of the final grade, the other 50% is achieved by building a project along the course.
Description of students' engagement with Telegram: The educator created a chat group for this course in Telegram and students were invited to join via a link. The chat group was created with the objective of providing students with fast communication with their educator, and the potential of including different bots. Students were invited to use the group to interact among each other and help others with their issues or questions. Different bots were integrated: a) a bot for reminding students about deadlines; b) a bot for moderating the chat group; c) a bot developed by the educator for automatically responding to FAQs (@faq_secodis_bot, freely available in Telegram); d) built-in polls to ask students their opinions and preferences along the course.
Affordances and challenges: The students' familiarity with the Telegram and the fact that they were already using it for an informal class group made them shift to the "official" chat group with their educator to ask questions and share solutions with other students, or share interesting information about the topics covered in this course. Engagement was around 100% in this chat group, and most students participated at least once asking questions, responding, sharing their thoughts or participating in the different polls during the course. Also, an important finding was that students were willing to help others by explaining their own solutions which favoured the best learning process of all. On the other hand, synchronous communications are sometimes very demanding especially with large groups. In order to ease this burden, specific rules for the chat group were considered, e.g.: asking students not to post questions after 9 pm or during the weekends/holidays.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Overall opinion of educators participating in Pilot: After Pilot 3 results, the experience from educators with the new tools was very positive. It is important to highlight that this time, students were using Telegram and that the initial enrolment and their participation in the chats was very high. Also, educators considered it very relevant to add new bots with different functionalities that not only helped educators but also encouraged students, probably because of the novelty of the personal interaction. Educators also decided to keep using these tools after the pilot for their new courses.

5.1.3. Courses at the University of Leeds

For Pilot 4, we decided to trial with a different cohort of students. Previous pilots at Leeds University with taught module students had been largely unsuccessful, due to factors beyond our control. Leeds University IT service has mandated other systems for student-educator interaction and for eliciting course feedback from students, and we were strongly discouraged from offering alternatives to our students. Furthermore, existing students were already used to these systems, and were reluctant to use the alternative platforms we offered. So, we sought a different type of student body for Pilot 4. Leeds University runs a Summer School every summer, hosting several hundred undergraduates from other universities around the world for intensive 4-week courses. These visiting students were not standard Leeds students, and not mandated to use the Leeds IT infrastructure and required systems for interaction and feedback. Furthermore, they were not already familiar with the Leeds IT standard services for chat and feedback; so they should be open to trialling new systems Differ and Hubert. So, we decided to run Pilot 4 with students at the Leeds International Summer School in summer 2021.

Unfortunately, shortly before it was due to start, the Leeds International Summer School was CANCELLED due to the Covid crisis. We had to come up with a last-minute alternative student cohort for Pilot 4: we identified 3 groups of University applicants who were not yet students but had applied to become Leeds University students starting in late September 2021. So, we offered Differ communities for 3 groups of “pre-sessional” applicants: (1) around 600 applicants to the Lifelong Learning Centre for mature students; (2) around 700 applicants for the Plus Programme, for students from disadvantaged or “less represented” backgrounds; and (3) around 500 applicants for Computing BSc, MSc, PhD programmes due to start end September 2021.

Best case

Course name: Plus Programme

Number of students: approximately 700 applicants were offered Differ
Course description: Not a taught course, but a support programme for Leeds University applicants from disadvantaged or “less represented” backgrounds, to encourage them to take up their offer of study at Leeds University.
Description of students’ engagement with Differ and/or Hubert: Applicants were offered use of Differ as a platform to communicate with each other and with Plus Programme administrative staff.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Affordances and challenges: This was not an ideal cohort for several reasons: (1) unlike Leeds University registered students and the Summer School students, we had no control over applicants prior to registration and arrival at Leeds, so we could only suggest they try Differ; (2) applicants likely already used widely-used general chat platforms (Facebook messenger, whatsapp etc) and Differ would not be novel to them; (3) these are not “educator-led” classes, as the Plus Programme is not staffed by academic educators but by clerical staff; (4) on arrival at Leeds, all new students would be required to take up the standard Leeds IT systems so any take-up of Differ would be temporary.

Overall opinion of educators participating in Pilot 4

The PIlot 4 cohorts were not educator-led, but 3 different groups of applicants who had not yet arrived at Leeds University but were encouraged to take up their offered places. The administrators who oversee student recruitment of these 3 cohorts thought Differ might offer an additional incentive or attraction to follow through their offer of study. The Plus Programme administrators were sufficiently impressed by Differ to commit to its continued use in applicant support, beyond the end of the EDUBOTS project.

5.1.4. Courses at the University of Zagreb

At the Faculty of Organization and Informatics (UniZg) only educators who had previous experience with **Differ** and already did piloting in earlier phases of the project participated in Pilot 4. Because of that, no onboarding for educators was organised. Onboarding for students was conducted in the beginning of the semester. students were invited to join Differ via LMS or face-to-face classes where Differ functionalities were demonstrated briefly on some courses. In total, Differ was used on 6 courses, also as Hubert evaluation - once or twice in the semester. Students were not interested so much in filling in the surveys due to parallel evaluations on faculty or university level that were conducted.

Best case

Course name: Intelligent systems

Number of students: over 100
Course description: The course is elective in the winter semester of the 2nd year of the Master's Degree at the university graduate study Information Systems. The goal of the course is that students: create a theoretical knowledge of intelligent systems technology and their most important techniques, which should enable the competence of students in solving specific complex problems of practical application, but also research. In addition, emphasis is placed on acquiring competence for the practical application of theoretical knowledge to solve complex problems. Through the exercises, the most effective methods of application are introduced and practised, and the theory from the lectures is deepened. This is achieved by applying the highest quality software tools in the field of intelligent systems, using illustrative examples from practice

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

and applying complex techniques and methods, which are well described theoretically in lectures and present in software tools. Teaching activities consist of the lectures and practical laboratory exercises. Students were mostly full-time, interested in the course topics since the course is elective.

Description of students' engagement with Differ and/or Hubert: Differ was introduced in the course as a part of a computer exercise. Students got the guidelines uploaded on the learning management system on how to join the community and use the app. An educator created in Differ topics for discussions with the students. Differ was used to further explain course announcements in LMS if needed. Students were actively participating in the discussions, mostly to get the answer to some exercise related problem or administrative questions. Most of the students used private messaging to directly ask the educator without other students seeing the message

Affordances and challenges: Motivate students to use Differ since they already used many applications.

Overall opinion of educators participating in Pilot 4

Educators who participated in Pilot 4, except one teaching assistant, were all who have already used Differ and Hubert in previous Pilots. Because of their previous positive experience with chatbots and advantages for students observed in previous pilots in online teaching, educators continued to use chatbots. New functionalities and improvements especially in Differ were additional encouragement to continuous use. For educators it was challenging to motivate students to use Differ because they have already used other communication apps and channels.

5.2. Interactions in Differ and Hubert

5.2.1. Analysis of system data in Differ

The Differ system data collected during Pilot 4 and the community categorisation were as described in section 2.4.1 Analysis of system data in Differ above, with additional events relating to users' interactions with the Bo chatbot also captured.

The number of communities created and volume of users ever active in Pilot 4 are shown below.

Community Type	Pilot 4 Communities	Users Ever Active by Differ Community		
		Minimum	Maximum	Mean
Course	8	0	40	14.6
Programme	9	3	45	15.9
Department	1	102	102	102.0
Multi-Disciplinary	16	0	145	44.1
Total	**34**	**0**	**145**	**31.9**

Table 9. Number of communities and active users in pilot 4

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

The average number of active users in each community across the four pilot phases is shown below.

Pilot Phase	Users Ever Active by Differ Community		
	Minimum	Maximum	Mean
Pilot 1	2	99	30.3
Pilot 2	0	87	18.0
Pilot 3	1	38	15.4
Pilot 4	0	145	31.9

Table 10. Average number of active in all pilots

The number of active users is somewhat difficult to interpret in the absence of data on the number of users invited to each community, but these data do reflect the observations of consortium members that some Differ communities were quite active while others failed to obtain much momentum.

5.2.2. Analysis of data from the course evaluations in Hubert

Two types of conversation templates were employed by project partners. **Course evaluation (CE)** questions were aimed at gathering feedback from students regarding the teaching of a given course or module, at or near the end of that course. **Self-reflection (SR)** templates, introduced during Pilot 3, were used at the beginning of, during, and at the end of some courses to encourage participants to think about their own learning. In addition to these core questions, each template used during the project also included one or more **meta-evaluation** questions to gather feedback on the experience of using Hubert itself. Many also captured some demographic data such as age and gender.

In a number of cases, users failed to reach the end of a conversation with the Hubert agent. While many of these incomplete interactions will still have provided potentially useful information, ones with no or extremely limited user responses are unlikely to be valuable for some analyses. To facilitate their (consistent) exclusion, the following definition of an abandoned conversation was applied, and further datasets generated which excluded such interactions:

*An **abandoned conversation** is one in which the user did not provide a response to any of the 'core' questions – the 'stop', 'start' and 'continue' questions for course evaluations, or the reflection prompts for self-reflection templates.*

The overall volumes of Hubert transcript data available are summarised in Table 11.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

		Raw Data		Cleansed Data		Final Data (Complete Conversations)	
Pilot	**Activity Type**	Files	Conversations	Conversations	Abandon Rate	Conversations	Questions
1	Course Evaluation	19	217	205	15.1%	174	1,324
2	Course Evaluation	13	170	165	15.8%	139	1,454
3	Course Evaluation	15	176	176	19.3%	142	1,407
3	Self-Reflection	7	43	43	14.0%	37	333
4	Course Evaluation	7	82	82	6.1%	77	757
4	Self-Reflection	6	74	74	10.8%	66	739

Table 11. Volumes of Hubert transcript data generated during the project

The duration of users' interactions with Hubert varied significantly, as shown below.

Duration	**Pilot 1**	**Pilot 2**	**Pilot 3**		**Pilot 4**	
	CE	CE	CE	SR	CE	SR
≤ 1m	6	9	10	0	5	3
1-2m	6	6	6	1	5	1
2-3m	2	10	11	1	9	2
3-5m	27	31	28	8	19	12
5-7.5m	43	26	29	5	22	17
7.5-10m	25	25	16	7	7	8
10-20m	50	25	33	9	9	15
20-100m	13	7	7	6	1	7
> 100m	2	0	2	0	0	1

Table 12. Duration distribution of Hubert conversations by Pilot phase and activity type.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

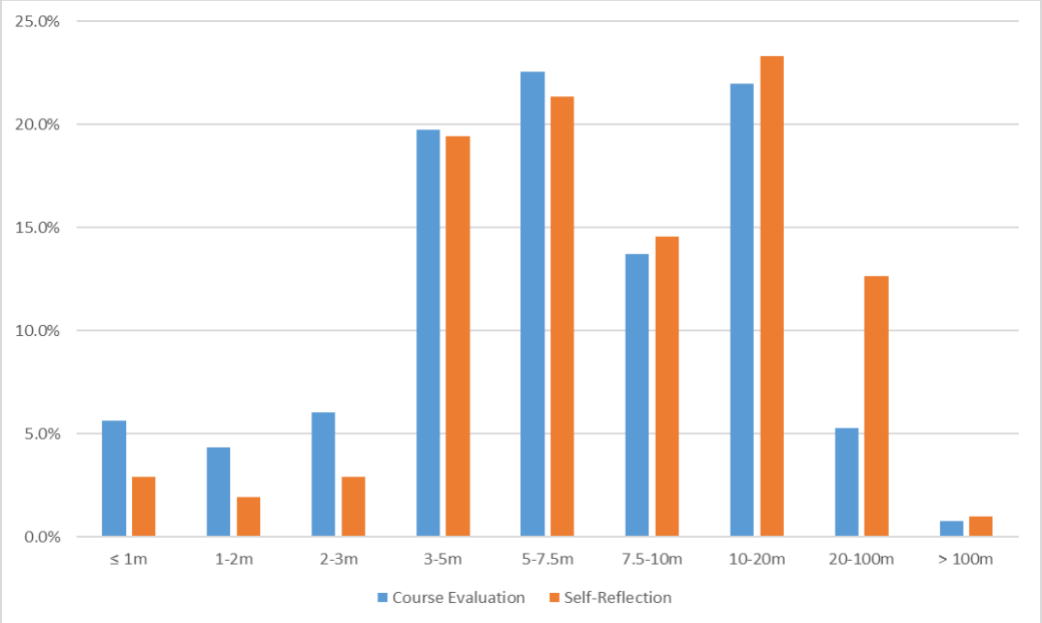


Figure 68. Duration distribution of (non-abandoned) Hubert conversations by activity type

There is some variation between the pilot phases, with Pilot 1 conversations tending to have somewhat longer durations. This may be due to changes to the software during the course of the project, including additional language support which may have allowed a proportion of users to provide similar feedback more quickly by working in their native tongue. There are also instances where very long apparent durations are calculated, which appear to be the result of users returning to an interaction which was previously interrupted.

Self-reflection conversations tended to take longer than course evaluations. The templates used were similar in terms of the number of questions included, so this observation may be due to the more thoughtful nature of the self-reflection exercise.

The Hubert chatbot's primary aim during course evaluations is to improve the overall quality of responses obtained from students when compared with more traditional feedback-gathering methods such as questionnaires. While the EDUBOTS project was not able to obtain directly-comparable data gathered using such approaches[3], manual inspection of the transcripts and annotation of constructive feedback do provide indications that Hubert is achieving at least some success in this regard; 19 of the 96 question-answer pairs in the Pilot 2 transcripts which contained constructive suggestions were obtained after an initial response such as "Nothing

[3] Students on some courses were asked to provide feedback using both traditional questionnaires and via Hubert, but the questionnaire response rate was extremely low, and anonymisation means that it is not possible to understand how much overlap exists between the respondents. Data was also not readily available for previous years, and such a comparison would in any case have been significantly complicated by the impacts of COVID-19 on teaching and delivery methods during the project.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

comes to mind right now" was followed up by the chatbot ("Surely there must be something which can be improved?").

5.2.3. Hubert meta-evaluation

As noted above, each Hubert conversation template used in the project included one or more questions which asked users to evaluate their experience of using the Hubert chatbot tool.

For Pilot 1, the question asked was "How did you like this type of evaluation compared to a regular survey?", with a free-format response. The bot then followed up by asking for a numeric rating, e.g. "So on a scale from 1-10, would you give it a 7?"[4].

For subsequent pilot phases, the question set was expanded to cover specific aspects of the bot's functionality, and the free-format responses were replaced with a fixed Likert scale, ranging from 1-5 for enjoyment and from 1-7 for the capability and ease questions:

- "How much did you enjoy using me (Hubert), from 1 to 5?"
- "Do my capabilities meet your requirements?"
- "Am I easy to use?"
- The scores obtained are summarised below.

Pilot Phase	Question Type	Responses	Mean Response	Scale Maximum
Pilot 1	Overall Evaluation	76	8.18	10
Pilot 2	Capability	93	4.06	7
Pilot 2	Ease of Use	93	6.00	7
Pilot 2	Enjoyment	119	4.38	5
Pilot 3	Capability	73	4.36	7
Pilot 3	Ease of Use	73	5.82	7
Pilot 3	Enjoyment	149	4.32	5
Pilot 4	Capability	119	4.39	7
Pilot 4	Ease of Use	119	5.00	7
Pilot 4	Enjoyment	122	4.30	5

Respondents on average scored Hubert quite highly for ease of use and enjoyability, but seem to have been slightly less enthusiastic about the bot's capabilities. Possible reasons for this were explored through the user surveys administered throughout the project.

[4] The suggested rating was based on sentiment analysis tools applied to the preceding user response.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

5.3. Results of the surveys and interviews about the chatbots

5.3.1. Differ and learning process

Respondents of the Pilot 4 study were 40 students: 36 of them were from Faculty of Organization and Informatics, and 4 of them from Universidad de Granada. Participants evaluated the role of chatbot-based communication platform Differ for academic purposes in the winter semester of 2021/2022. academic year. Data was gathered on five courses at undergraduate level of study. Regarding demographic characteristics of the sample, 70 % male and 30 % female evaluated Differ.

Interconnection of successful course completion, engagement and Differ use

Regarding frequency of use, most of the students used Differ once a month. Only two students used Differ several times a month, only four students once a week, and three students several times a week.

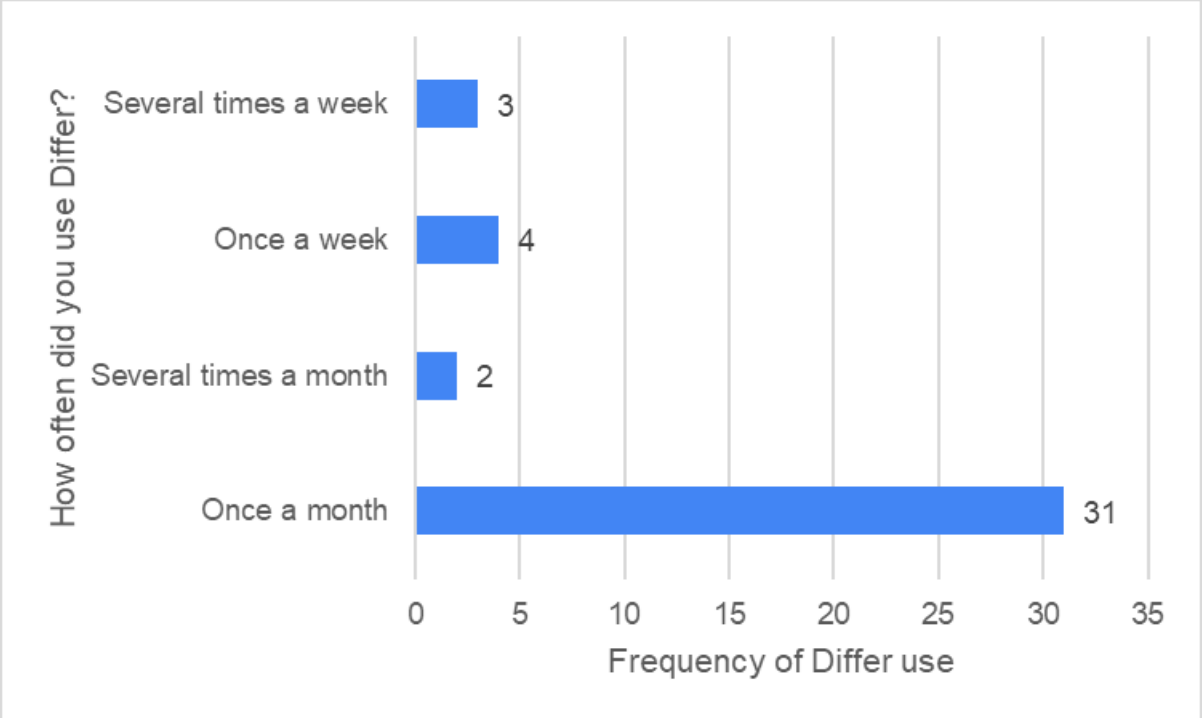


Figure 69. Frequency of Differ use in a pilot study, n=40

From analysis of students' responses to the two items related to **course performance and grade achievement,** following results were obtained:

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

- Most of the students are **indifferent** regarding Differ helping them in achieving better grades.
- 42% of the students considered Differ as a tool that did not help them for successful course completion.

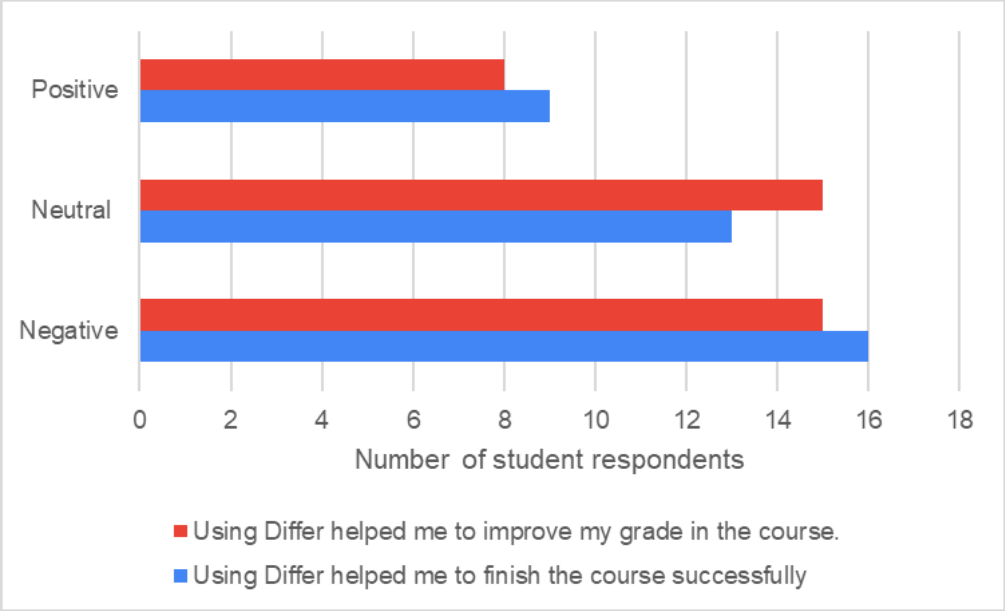


Figure 70. Student responses related to course performance, n=40

To examine the relationship between the frequency of use of the Differ application and student academic success, Spearman correlation coefficient was calculated and results are shown in Table XY.

Learning engagement items	**Spearman's correlation coefficient**
Using Differ helped me to achieve a better grade.	.2269
Using Differ helped me in successfully completing the course.	**.3078****

**. Correlation is significant at the 0.01 level (2-tailed).

Table 13. Results of correlation analysis of frequency of using Differ and learning engagement items

We can conclude that in the case of this pilot study, more frequent use of Differ **was correlated with more successful course completion among students.**

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

The role of the Differ application in connecting students with colleagues and educators in the course

From the interpretation of student's answers from the survey following results were obtained:

- Differ did not have a positive impact on connecting students. Most of them think that Differ did not help them to connect with other students.
- Chatbot application was also not recognized as particularly useful for collaboration on the course-related activities.
- Furthermore, Differ also did not make an impact on student-educator connections.

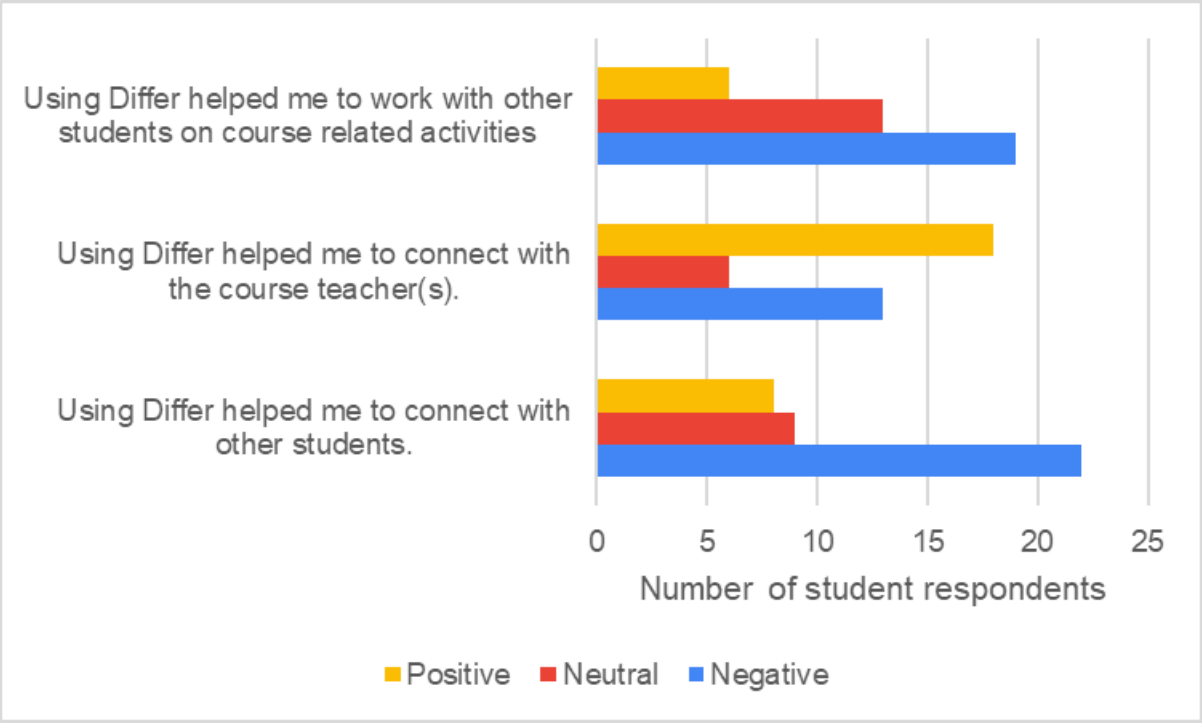


Figure 71. Differ's role in peer-to-peer and educator's connections

5.3.2. Differ and teaching process

In the Project as the part of Pilot 4 we again performed a survey about the use of Differ and Hubert with educators. The goal was to find out how educators use Differ and Hubert as well as how it helps them in the teaching process. Again, a small number of educators, only 11 (4 from Croatia, 4 from Cyprus nad 3 from The United Kingdom), participated in the survey. Among the educators 7 used Hubert, while 8 used Differ. Quantitative data will be presented only through descriptive statistics.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

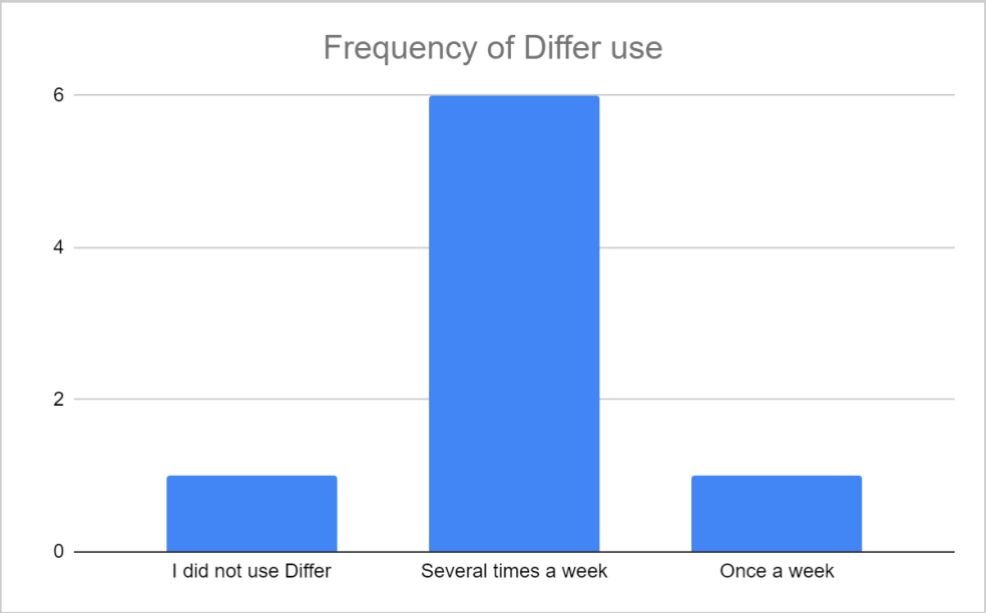


Figure 72. Frequency of Differ use

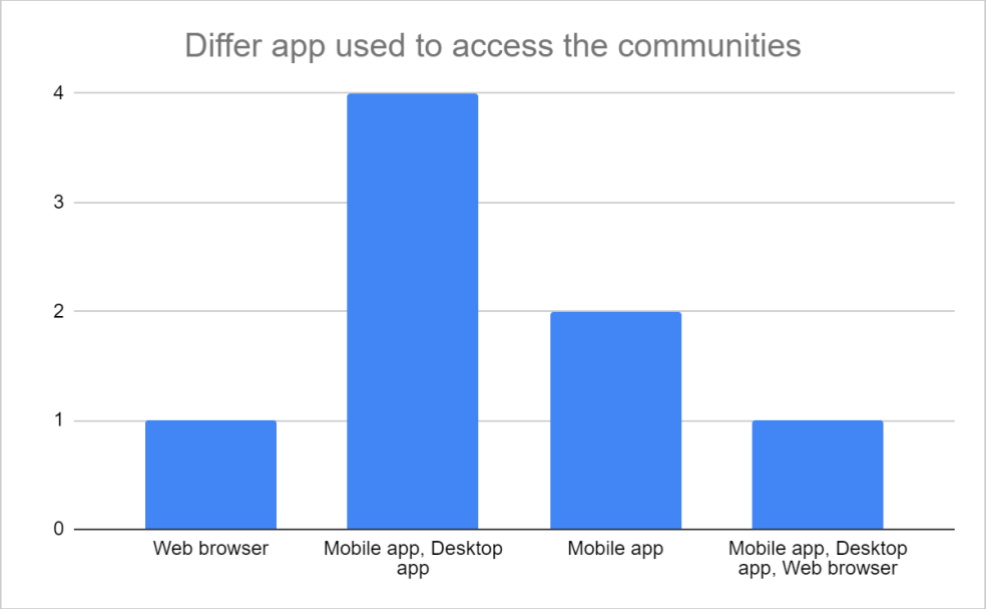


Figure 73. Differ app used to access communities

Educators mostly used a combination of desktop and mobile apps. This is a slight shift towards the combination of apps which are used, when compared to the previous pilots.

Three educators used Differ in a single course, two used it on two courses, twou used it on three courses and one used in more courses.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

First graph focuses on the **educator's perception of Differ's impact on students**. Results indicate mixed attitudes in most statements except for the inappropriate use of communication style, where educators dominantly disagreed, thus their perception is that students used appropriate communication style.

Second graph focuses on nine statements related to the **Differ's impact on the teaching process**. In most of the statements results are again mixed. Survey was conducted on a too low number of participants to draw any statistically justifiable conclusions.

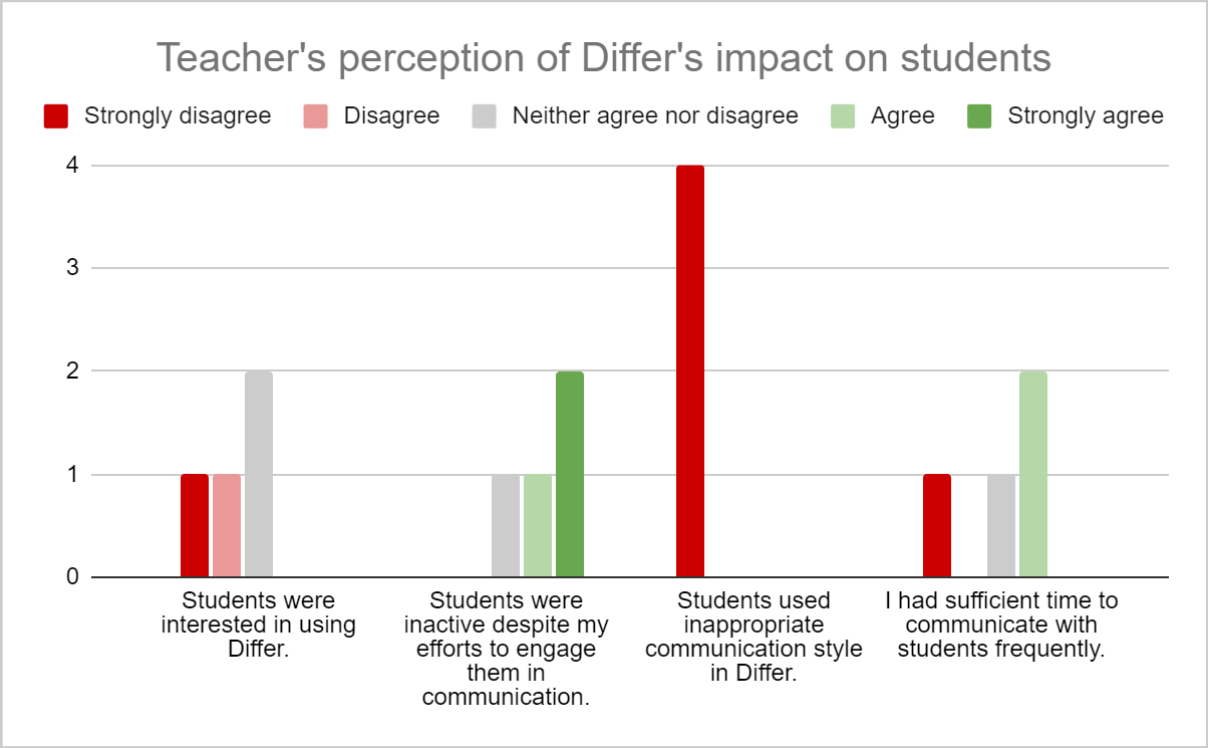


Figure 74. Teachers' perception of DIffer's impact on students

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

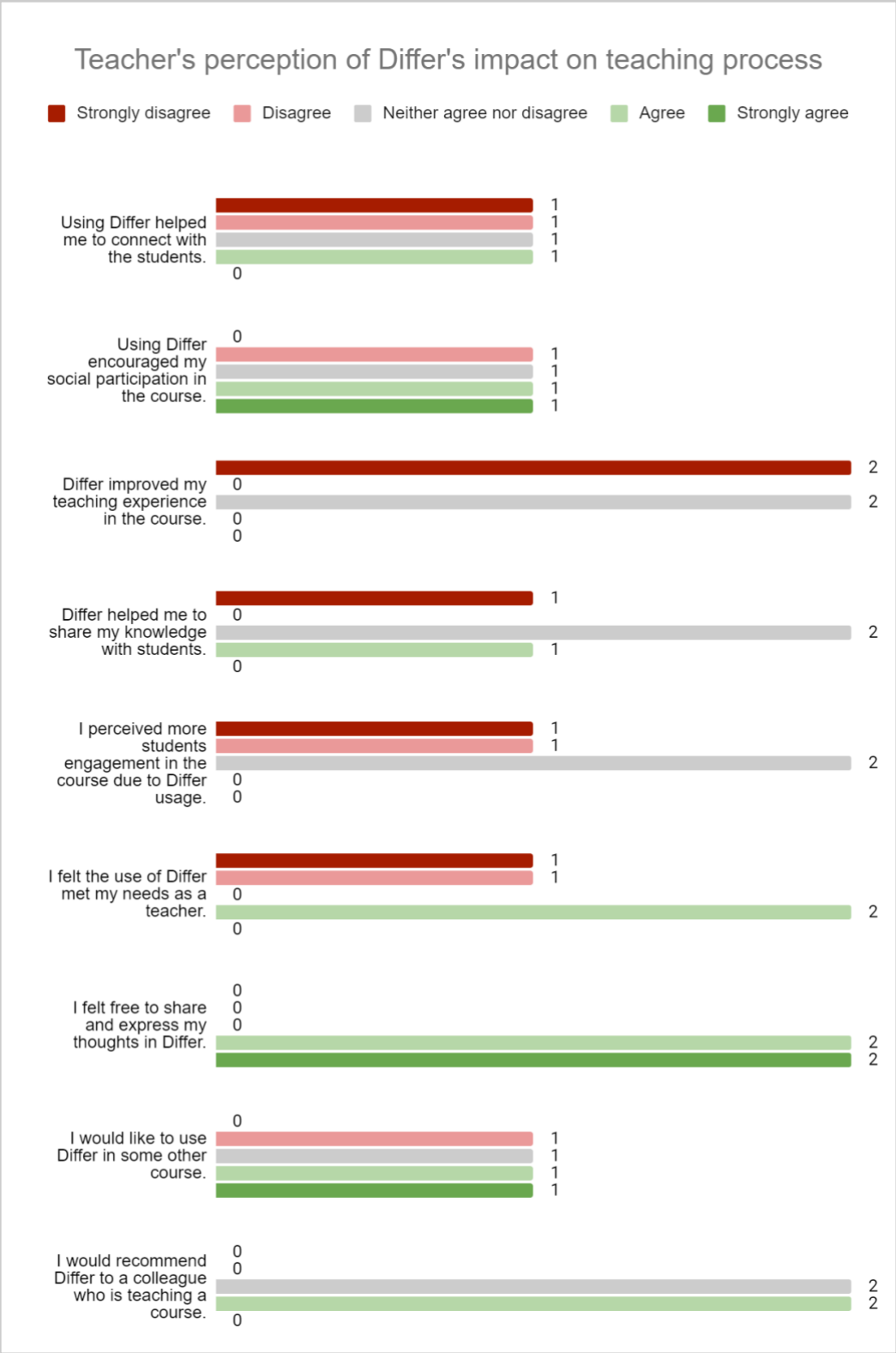


Figure 75. Teachers' perception of Differ's impact on teaching process

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Two educators used Hubert in one course, three educators used it in two courses and two educators used it in three courses. Among the educators four used Hubert once in the courses, while three used it two times in the courses.

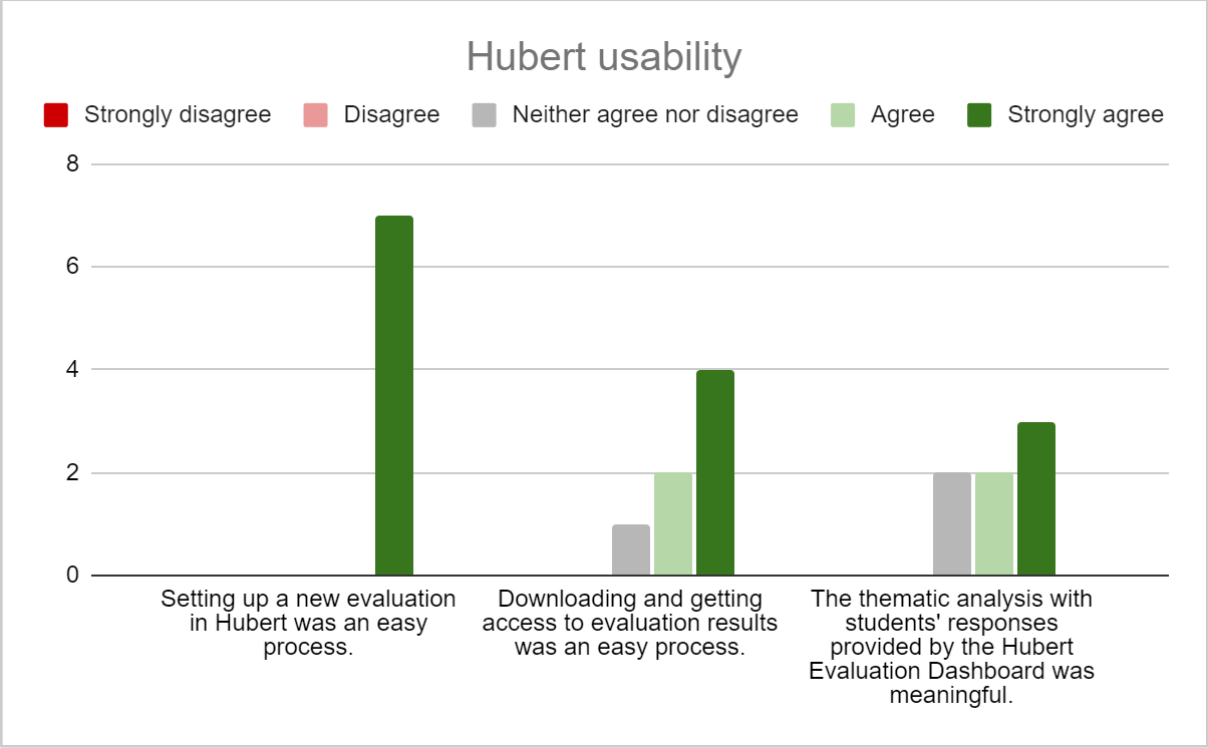


Figure 76. Hubert usability

Regarding the Hubert usability experience, results indicate dominantly positive aspects related to the setup process, download and access to the results, and thematic analysis of the responses.

Differ and Hubert were perceived as tools with minor impact on the proportion of students who successfully completed the course. Out of six educators, two answered that the tools had a positive impact, while four answered that the tools did not have an impact. Educators who answered positively provided argumentation related to the Differ and Hubert communication capabilities related to the student engagement and feedback. Educators who answered negatively provided arguments which are based on infrequent use of the tools.

Educators joined Differ equally by invitation from the EDUBOTS web page and by invitation from their colleagues.

Educators were also equally split among those who used Differ in english because they have chosen so and those who used it in English because they did not notice the option to change the language.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

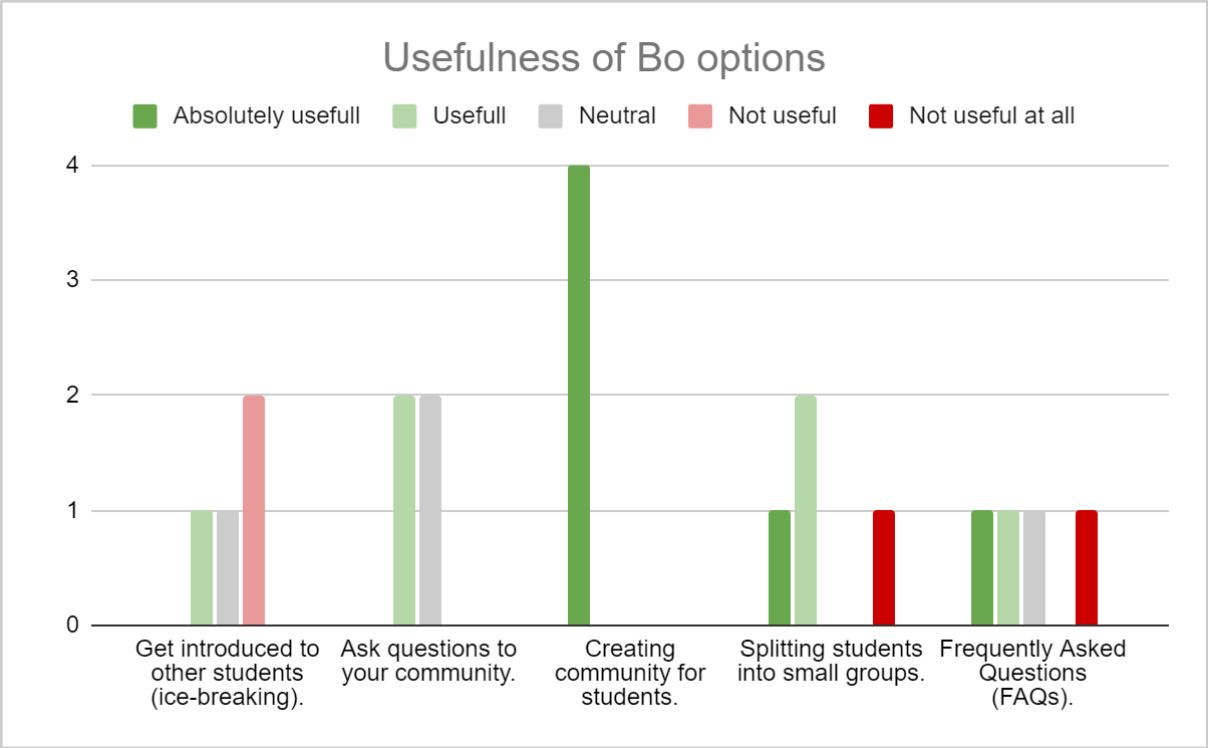


Figure 77. Usefulness of Bo options

5.3.3. Usability and User Experience of BO in Differ

In Pilot 4, we decided to unburden the students with a lot of surveys and put focus only on chatbot BO. Thus, we measured the usability of BO only with the Chatbot Usability Questionnaire (CUQ). Authors of the questionnaire CUQ suggest that CUQ and User Experience Questionnaire (UEQ) measure similar aspects of chatbot usability and UX, identified by multiple regression (Holmes et al 2019).

In addition, the CUQ is designed to be comparable to SUS with addition of statements related to chatbots, and includes 16 items. Participants' levels of agreement with sixteen statements are ranked out of five, from "Strongly Disagree", to "Strongly Agree" (Holmes et al 2019).

However, the questionnaire was adapted in Pilot 4 to more precisely measure usefulness, appropriateness and informativeness, so the statement "Chatbot responses were useful, appropriate and informative." was separated into three statements.

In Pilot 4, 39 respondents filled in the survey, 35 from the University of Zagreb, Faculty of Organization and informatics, and 4 from the University of Granada. The mean answers are provided in Figure 78. The greatest scores were received for the statements "The chatbot was welcoming during initial setup.", "The chatbot was very easy to use." and "The chatbot was very easy to navigate."

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

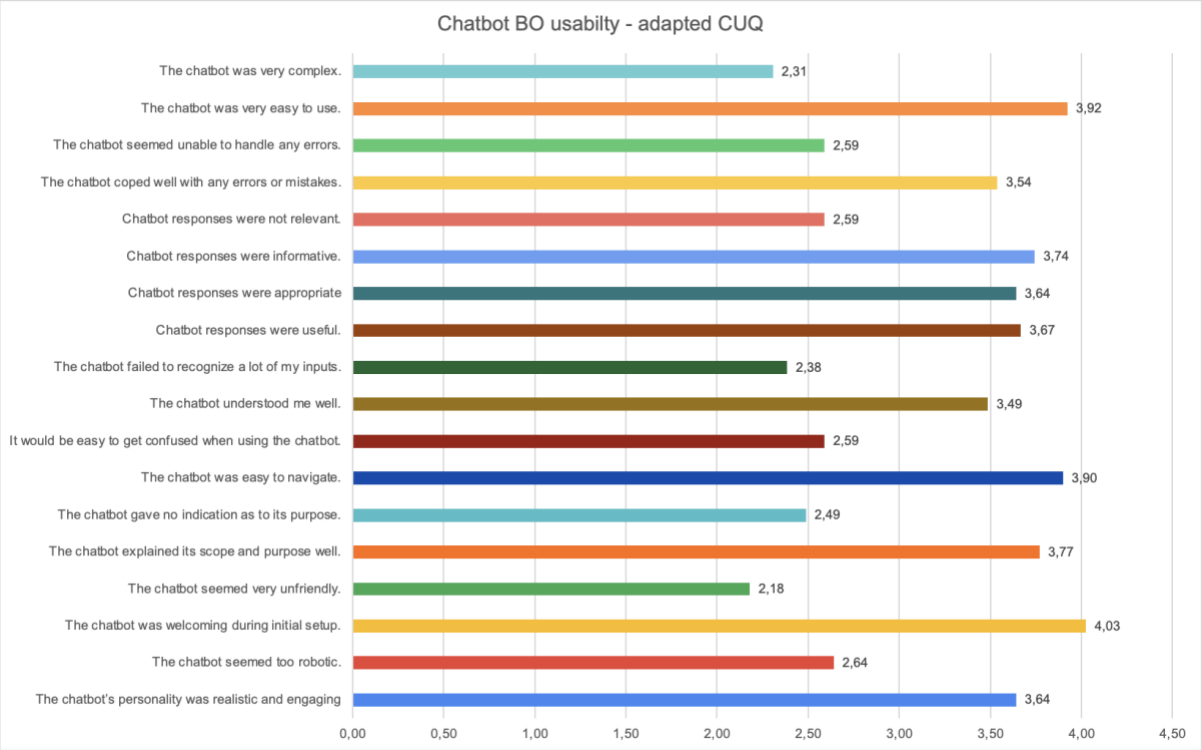


Figure 78. Usability of BO evaluated with CUQ (Pilot 4)

Since similar values in three separated statements were identified (mean 3,67 for useful, 3,64 for appropriate, and 3,74 informative), we decided to calculate the mean of them and then compare it with the results from previous pilots.

From the graph, we could see that values in Pilot 3 and 4 are pretty similar, and that respondents in both pilots perceived BO in Differ quite uniformly. Additional statistical comparison would give a precise answer if there is difference between the respondents in Pilot 3 and 4.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

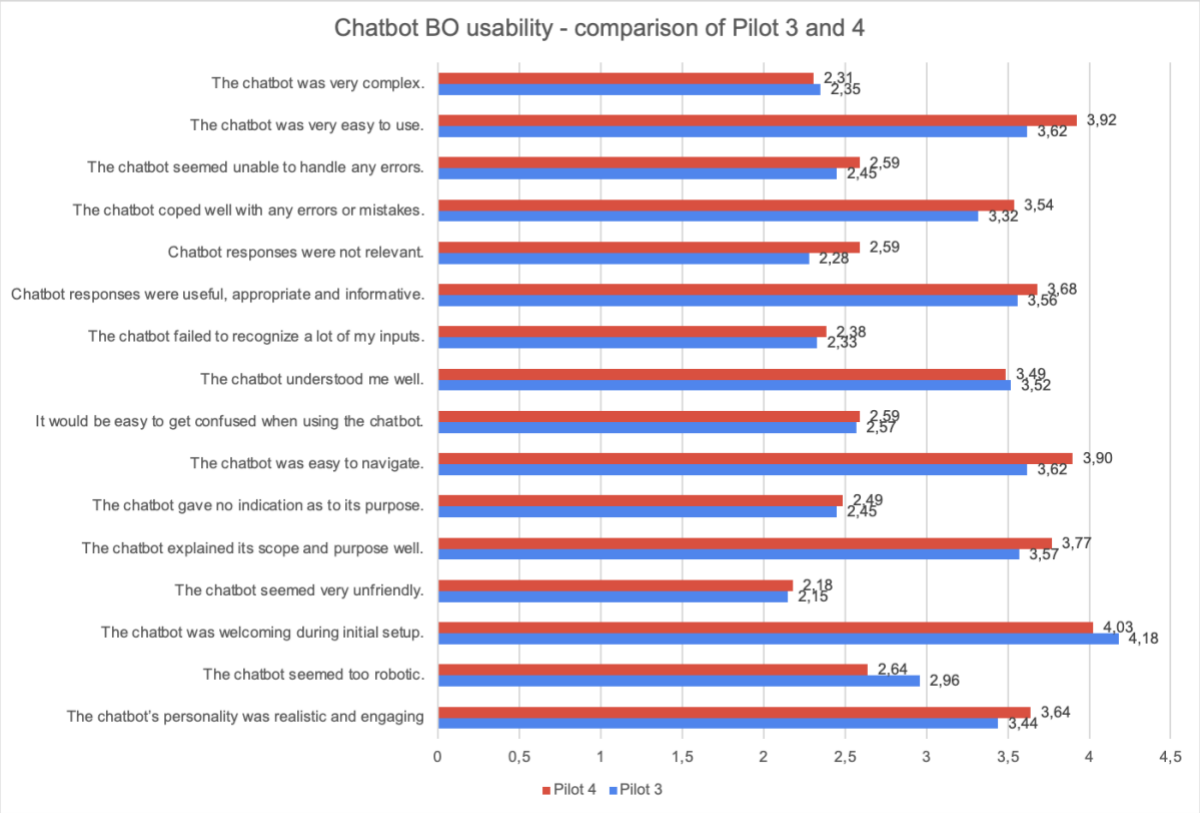


Figure 79. Usability of BO evaluated with CUQ (Pilot 3 and 4)

5.3.4. Usability of Hubert

In Pilot 4, evaluation of usability and user experience of the chatbot Hubert was performed with questions from the questionnaire UMUX-Lite (Lewis et al, 2013) and adapted Chatbot Usability Questionnaire, CUQ (Holmes et al 2019).

Usability of Hubert evaluated with UMUX-LITE

Two questions from the UMUX-LITE were integrated at the end of the conversation with Hubert with which students evaluated the course or made self-reflection related to the topics in the course. With the question “Am I easy to use?” perceived ease-of-use was measured, while with the question “Do my capabilities meet your requirements?” perceived usefulness was measured, both with seven-point scale items.

Overall, 119 students had responded to the two items related to usability, 27 from the Cyprus University of Technology and 92 from the University of Zagreb, Faculty of Organization and Informatics. Data provided still indicated that some students might have been confused with the reverse scale (1 - Strongly Agree to 7 - Strongly Disagree). Ten responses were identified as outliers, so calculation was based on 109 responses. When calculating scores, the scale used in

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

the conversation was reversed to provide correct calculations (1 - Strongly Disagree to 7 - Strongly Agree).

Based on those responses, perceived usability of Hubert in Pilot 4 was slightly lower than in previous two pilots: 68,4 on the scale from 0-100, which is average or good score.

Perceived ease of use was rated lower than in previous pilots (mean 5,4, st.dev. 2,3), but Hubert capabilities (perceived usefulness) was little bit higher than in previous pilots (mean 4,83, st.dev. 2,1) as it can be seen in Table 14.

		Responses	Mean	St. Deviation	Scale Maximum
Pilot 2	Overall Evaluation	87	**71,07**	23,46	100
	Ease of Use	87	6,33	1,48	7
	Capability	87	4,20	2,24	7
Pilot 3	Overall Evaluation	69	**71,86**	24,00	100
	Ease of Use	69	6,10	1,50	7
	Capability	69	5,05	2,00	7
Pilot 4	Overall Evaluation	109	**68,43**	26,67	100
	Ease of Use	109	5,39	2,28	7
	Capability	109	4,83	2,09	7

Table 14. Perceived usability of Hubert across the pilots (UMUX-LITE)

Across all pilots we see that mean responses are rather similar, so further statistical comparisons were not conducted. The reason for similar means might lay in the fact that Hubert was not upgraded in terms of functionality or user interface design that really affected the end user.

Usability of Hubert evaluated with adapted CUQ

In addition to UMUX-LITE, usability of Hubert in Pilot 4 was further explored with the Chatbot Usability Questionnaire (CUQ), in a separate online survey implemented in the Google Forms.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

The CUQ includes 16 statements and participants' levels of agreement are ranked out of five, from "Strongly Disagree", to "Strongly Agree" (Holmes et al 2019).

The original questionnaire from (Holmes et al 2019) was adapted in Pilot 4 to more precisely measure usefulness, appropriateness and informativeness, so the statement "Chatbot responses were useful, appropriate and informative." was separated into three statements, each having one attribute. Adapted questionnaire had 18 questions.

However, a partner from Cyprus decided to further shorten the questionnaire by omitting opposite statements like "The chatbot explained its scope and purpose well." and "The chatbot gave no indication as to its purpose.", and left only positive statements, which resulted in 8 statements.

In Pilot 4, in total 94 respondents filled-in the survey, 68 from the University of Zagreb, Faculty of Organization and informatics, and 26 from universities in Cyprus. The mean answers are provided in Figure 80. The greatest scores were received for the statements "The chatbot was welcoming during initial setup.", "The chatbot was very easy to use." and "The chatbot was very easy to navigate."

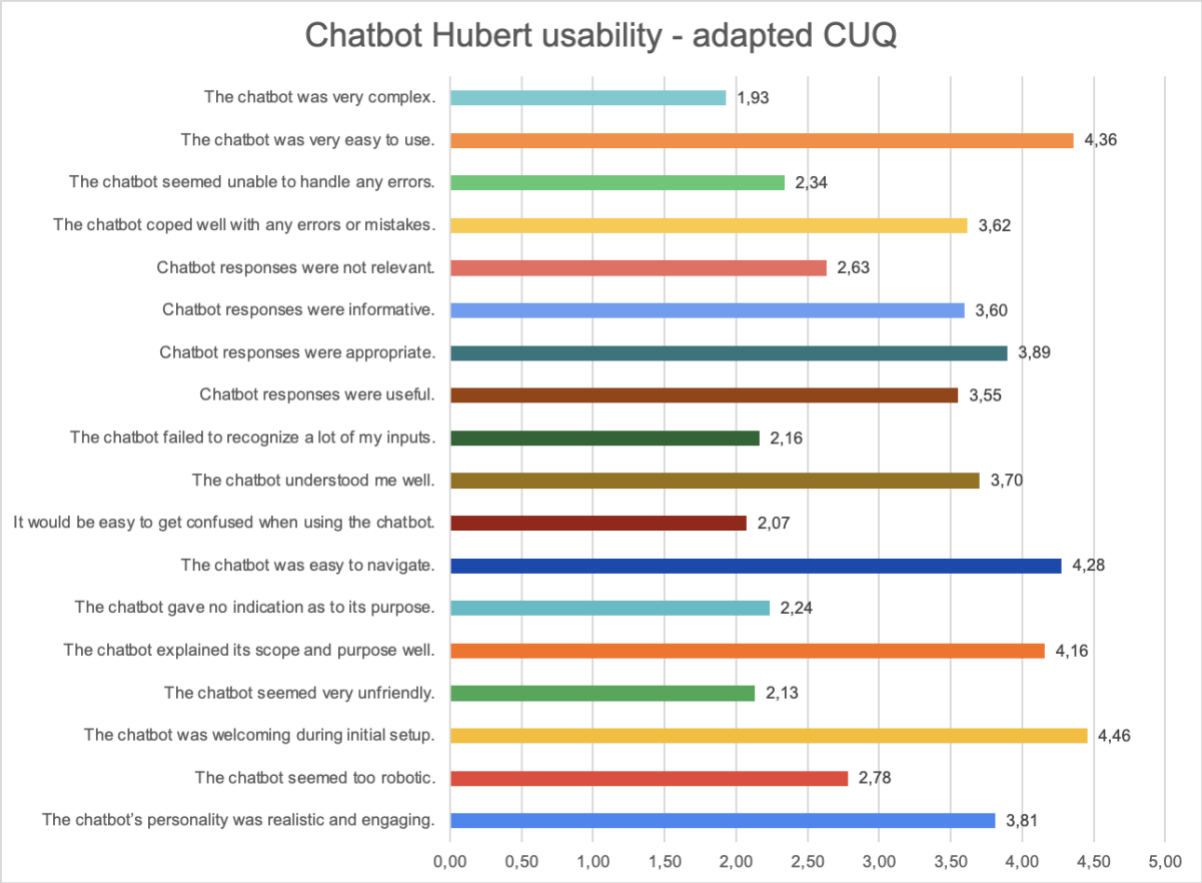


Figure 80. Perceived usability of Hubert across the pilots (adapted CUQ)

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Several students provided additional comments about Hubert - most of them were very positive:

> *Everything is great, interesting.*
>
> *Interesting program, I like it very much.*
>
> *Maybe that his answers are not so uniform :)*
>
> *Impressive chatbot.*
>
> *Polls at the end of the interview in which we have to give an assessment of Hubert, list the questions too quickly, they can't be read*
>
> *The UI is a bit weird at times and the Croatian-English combination has some flaws.*

Some of the students commented on the scale of UMUX-LITE questions, which gave us proof that students might have incorrectly scored Hubert due to misinterpretation of the reverse scale (1 - Strongly Agree, 7 - Strongly Disagree).

> *Regarding the rating, in the end he says that the scale is from 1 to 7, for 1 I completely agree, and 7 I completely disagree, and in fact it should be vice-versa, so it needs to be corrected.*
>
> *Question: "Do my capabilities meet your requirements?" gives a scale of 1 (agree) - 7 (disagree) and then the interpretation of the answer is reversed from that of the question*
>
> *Hubert asked me to evaluate him on the scale from 1 to 7. I gave a 7 (the most positive response) and he answered "I should quit my job :-(". "*
>
> *In the end, when he asked for feedback, he asked if it was easy to use and that the number 1 signifies complete agreement with this thesis, but it "saddened" him.*
>
> *So, to put it bluntly, he is a legend, it was nice to talk to him, but there was a small mistake at the end of the conversation with him. He asked me to rate him, but I put it as it was written for the highest grade, so he reacted sadly.*

In addition, we were interested to see how respondents' answers about perceived usability of Hubert (N=94) are compared to those of BO in Differ (N=39). The Figure 81 presents a comparison of means for Hubert and BO. Better positive scores slightly go in favour of Hubert in 14 statements out of 18 (77.8%).

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

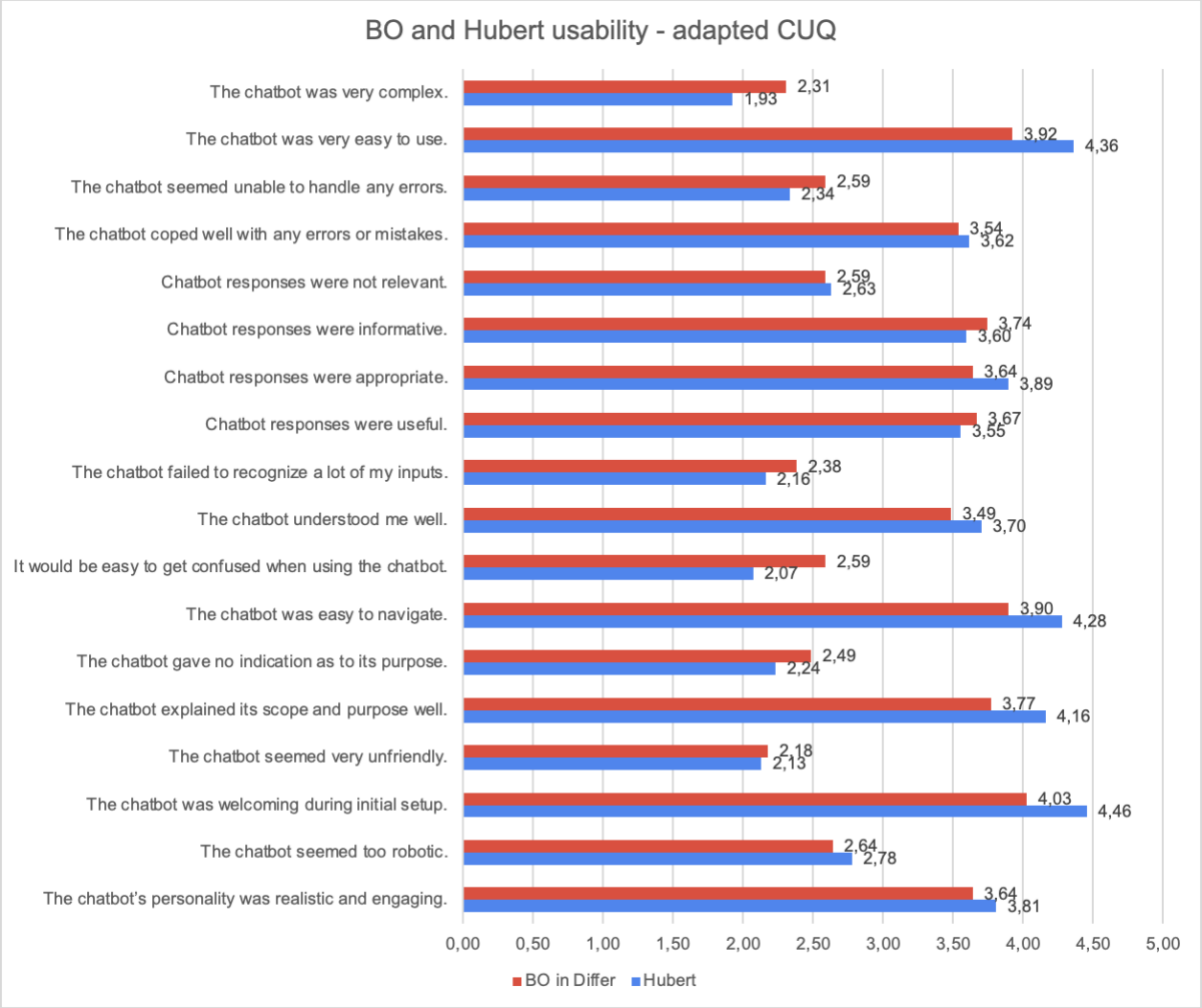


Figure 81. Perceived usability of BO and Hubert in Pilot 4 (adapted CUQ)

5.3.5. Results from the interviews with the students and educators

Analysis of the students' responses

Interviews with students were also conducted as a part of the Pilot 4. During courses, educators asked their students to participate in the interviews. Four students have agreed to participate in the interview.

Two students that have participated in the interview are on the undergraduate level (one is attending Information Systems and Business Systems and the other professional study programme Application of information technology in business). Two students are from graduate

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

study programmes - one attending Multimedia and graphics program and the other Information and software engineering. All students have encountered chatbots in their lives before Differ and Bo, mostly in the field of customer service.

Focus of the interview were: a) previous experience with chatbots and b) experience with BO, Differ and Hubert.

In the first part of the interview interviewers asked different questions related to positive and negative experiences with chatbots. One interviewee noted that their experiences were positive, one negative, and two of the students were neutral.

Usage of BO options among participants:

- Welcome to Differ – all participants used it.
- Get introduced to other Students – none of the participants used it.
- Ask questions to your Community – none of the participants used it.

BO left a good impression on all participants. One of the participants suggests implementing speech recognition.

The interviewees primarily used Differ for announcements related to the course (one participant), for communication with educators (two participants), for direct communication with colleagues (one participant).

Interaction with Hubert was interesting for participants. However, they see possibility for improvements in terms of explanations for questions. Their experience showed that Hubert only repeated the questions and could not give any explanations. Other improvements could be integration with other platforms such as Facebook Messenger, and use of Hubert as a search engine.

SWOT analysis of educators responses

To get a detailed view on the functionalities and experience of usage of Differ, BO and Hubert in various educational case scenarios, SWOT analysis was performed from responses of interviews with educators who participated in the Pilot 4. Participants consisted of educators from Croatia and Cyprus. SWOT Analysis is a useful framework for identifying strengths, weaknesses, opportunities, and threats of a chatbot application Differ. From the outcomes of this evaluation, we could identify main advantages and opportunities for other use case scenarios that could be reached in the future in higher education scenarios. On the other hand, recognized weaknesses require better developed functionalities that will insure technical stability on all available operational systems and devices (desktop, mobile, browser version). Important challenges are also listed and should be tackled down to minimize risks that come from external environments including other alternative technical and communication solutions that are commonly used in higher education.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

From the interviews, valuable feedback was collected in the mean of technical, usability and user experience improvements that will be taken into consideration before releasing new versions and functionalities of Differ, Bo and Hubert

Strengths	Weaknesses
BO and Differ ● The interface itself is very clear and very intuitive. ● It is easy to send group and private messages. ● Good for communication. ● Easy to use and understand. ● Friendly tone of the conversation. ● Creating a community option is very useful. ● Differ helped to achieve a sense of closeness during online teaching through interaction with students in small groups. **Hubert** ● Receiving students' feedback in an informal way. ● Accessing Hubert raw data which provided feedback on teaching. ● Anonymity of Hubert made students freer to be honest. ● Anonymity of Hubert made the interaction easier. ● Available templates which enable fast creation of questionnaires. ● Fast and easy access to the results.	**BO and Differ** ● Inability to share attachments as e.g. Facebook Messenger. **Hubert** ● There is no intelligence in answers and explanations. ● Hubert thematic analysis is a weak point - sentiment analysis is not sufficient. Hubert marks the answer as negative if the student had nothing else to add. ● The methodology for interpreting the results is not clear.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Opportunities	Threats
BO and Differ ● Possibility for BO to answer the FAQs, not just help educators to set up FAQs. **Hubert** ● Implement more use cases. ● Conduct small quizzes on specific topics with Hubert as a way to assess the students. ● Availability in other languages. ● Option to create your own questions or at least to have a pool of questions to customise templates. ● Use Hubert results in teaching process improvement. ● Implement machine learning algorithms.	**BO and Differ** ● Availability of similar communication applications with better and more mature platforms. **Hubert** ● Existing learning tools and communication channels that universities use for years.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

6. Best practices in using chatbots – research papers

Fourteen research papers presented at the conferences or published / accepted for publishing in journals within the scope of EDUBOTS project are presented here with their abstracts:

1. **Tsivitanidou, O., & Ioannou, A.** Users' needs assessment for chatbots' use in Higher Education, 31st Central European Conference on Information and Intelligent Systems, - CECIIS 2020, http://archive.ceciis.foi.hr/app/public/conferences/2020/Proceedings/ELA/ELA3.pdf

   Abstract. Higher education comprises an important field for the application of chatbots, especially for large-scale use. This paper reports on a needs assessment that was conducted with higher education users (i.e., educators and students) for examining their needs and expectations on chatbots' integration in educational settings. The study was conducted in the context of a research project that includes a series of iterative pilot studies of the use of chatbots in higher education. We report on findings from one of the pilots based on data from semi-structured online interviews with higher-education students and educators. A thematic analysis of the interview data resulted in different themes of needs that users have in education. The outcomes of this study indicate that higher education users need technological solutions that can support content delivery, formative assessment implementation with the provision of qualitative feedback, research tasks processing and social bonding facilitation. Those findings, along with interviewees' suggestions on functionalities and features that chatbots should have, provide guidelines and recommendations for the design, development, and implementation of different scenarios of the use of chatbots in higher education.

2. **Merelo-Guervós, J.J., Mora, A.M., Castillo, P.A.** Using chatbots in higher-education classrooms: Expected benefits in an European pilot experience, IV International Virtual Conference on Educational Research and Innovation – CIVINEDU 2020, https://www.researchgate.net/publication/344992161_Using_chatbots_in_higher-education_classrooms_Expected_benefits_in_an_European_pilot_experience

   Abstract. Chatbots are conversational programs that simulate speaking -in a human-like fashion- with a person. Nowadays they are able to answer almost any speaker's question fluently, and even ask other questions to the human. Chatbots have become an incredibly useful tool in many domains and applications, being very famous the assistants created by Google, Apple (Siri) or Microsoft (Cortana), which are also able to interact with the human using the speech. Even if they have been used in a lot of areas, there are still very few applications of chatbots in education, and almost none in higher-education; due to the lack of knowledge regarding how to implement and train the chatbots for specific scenarios in this domain. The main goal of the proposal is to fill the existing gap between educators and students due to the usual existing high ratio, which makes it very difficult to achieve a

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

desirable formative assessment as well as providing personalised feedback. To address this challenge, this paper comments a tentative approach using two different types of chatbots in order to increase the feedback to the educator, to have a better student follow-up, and to reduce the dropout rates (mainly in the first academic years).

3. **Čižmešija, A., Hajdin, G. & Oreški, D.** (2021) Using chatbot for course evaluation in higher evaluation. In: INTED2021 Proceedings doi:10.21125/inted.2021.0343. https://library.iated.org/view/CIZMESIJA2021USI

   Abstract. Growing potential of chatbots is getting more attention in education, both in application and in the research area. Using chatbot technology as a new channel of collecting feedback from students is an innovative approach that is less time and resource-demanding for both sides included in evaluation: educators and students. Main purpose of this study is to present a result summary of the students' reflection upon courses collected in an innovative way – using the educational chatbot Hubert. Hubert's features include dynamic conversation, personal experience and better interpretation of student feedback. Built on the basis of user feedback, new features include a more dynamic chat conversation that allows for a more personalized experience and better comprehension of students' input. Hubert's main advantage is engaging students and customizing flow of conversation about course perception and asking follow-up questions for additional explanation. The study was conducted at three courses at the University of Zagreb, Faculty of Organization and Informatics in 2020. Number of students involved in study was n=69. Thematic analysis was conducted to reveal patterns and themes within a data set of student-chatbot conversations. Following Braun and Clark's six-step recursive framework, main themes and categories were found upon positive and negative aspects of the course, suggestions for improvements and students' personal opinion. Since course evaluation was conducted in full swing of COVID-19 pandemic, some of the themes are related to this topic. In conclusion, guidelines and recommendations for implementing chatbots as a form of course evaluation in higher education are highlighted.

4. **Čižmešija, A., Horvat, A. & Plantak Vukovac, D.** (2021) Improving student engagement and course completion using chatbot application. In: INTED2021 Proceedings Online Conference doi:10.21125/inted.2021.1697. https://library.iated.org/view/CIZMESIJA2021IMP

   Abstract. Fostering students' engagement in the learning process is crucial for their academic success and can impact their achievement, course completion and skill development. This paper explores the potential of chatbots for improving student engagement and course completion. For this purpose, we employed a communication platform with chatbot functionalities called Differ in five courses at the Faculty of Organization and Informatics during the summer semester of 2019/2020 academic year. Educators that tested the platform used different strategies to engage their students in Differ, which resulted in various amounts of student's interactions with the educators or among the students in the course. At the end of the semester, students received a survey

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

in which they were asked to evaluate their learning process and engagement using Differ. Overall, 76 students (n=76) responded to the online survey voluntarily and anonymously. For 84,21% of respondents, usage of Differ has improved communication with the course educators. Among participants, Differ has been recognized as a valuable tool for improving course performance which was confirmed by calculating Spearman's rho correlation coefficient. However, even though the chabot application gained a positive impression because of establishing a more approachable and informal communication channel with the course educators, results show lack of peer-to-peer communication habits on the Differ platform for course activities. Future research should include more students from different universities to gain better insight into the usage and interactions within the chatbots in education, and explore the educator's perspective with best practices to use the chatbots in the course.

5. **Plantak Vukovac, D., Horvat, A. & Čižmešija, A.** (2021) Usability and User Experience of a Chat Application with Integrated Educational Chatbot Functionalities. In: Zaphiris, P. & Ioannou, A. (ed.) *Learning and Collaboration Technologies: Games and Virtual Environments for Learning. HCII 2021. Lecture Notes in Computer Science*. Cham, Switzerland, Springer International Publishing, pp. 216-229 doi:10.1007/978-3-030-77943-6_14.
https://link.springer.com/chapter/10.1007%2F978-3-030-77943-6_14

   Abstract. In recent years chatbots have found their use in education, especially in higher education (HE), but are they really serving a purpose and helping both students and faculty staff? Since bad experiences can discourage potential users, chatbots have to be carefully designed to be useful and to offer the best possible experience to its users. To explore this premise, in this paper we will tackle one aspect of chatbot design through evaluation of usability and user experience of a chat application Differ, which has integrated chatbot functionalities. For the duration of one semester at the Faculty of Organization and Informatics' students used Differ on several courses. Differ was introduced as a tool to facilitate communication related to courses and students' obligations, between students and educators by initiating communication through a Differ's chatbot called Buddy Bot. Differ was also seen as a tool to enhance students' connections and social life through creation of private student communities. After the end of semester, students' opinions about Differ were evaluated through means of UEQ and SUS questionnaires which are frequently used for user experience and usability evaluation. Results of the study showed that Differ's perceived usability and user experience are slightly above average, but with a lot of room for improvements, especially in the context of novelty, stimulation and chatbot functionalities.

6. **Abbas, N., Pickard, T., Atwell, E., Walker, A.** (2021) University Student Surveys Using Chatbots: Artificial Intelligence Conversational Agents. In: Zaphiris, P. & Ioannou, A. (ed.) *Learning and Collaboration Technologies: Games and Virtual Environments for Learning. HCII 2021. Lecture Notes in Computer Science*. Cham, Switzerland, Springer International

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Publishing, pp. 155-169. ISBN 978-3-030-77942-9. https://www.springerprofessional.de/en/university-student-surveys-using-chatbots-artificial-intelligenc/19325608

Abstract. Predefined web surveys are often used to collect course evaluations from students in higher education institutions. These institutions use the evaluations to adjust their courses' pedagogical standards and lecture style to cope with an increasingly uncertain and complex world. Many limitations to using web surveys have been reported such as low response rates and low-quality responses to open questions. To overcome these limitations, artificial intelligence conversational agents (CAs) or 'chatbots' are used to play the interviewer role, facilitating the enhancement of the quality of responses. This is accomplished by mimicking human-human conversations; by asking questions in a friendly, casual way and pursuing high-quality responses. This study aims to explore the opportunities and the obstacles of using CAs in collecting course evaluations in three European universities (UK, Spain and Croatia) and one Centre of excellence in Cyprus. The transcripts collected have been analyzed using statistical data analysis methods and qualitative data analysis techniques. Our findings reveal that the use of CAs in collecting course feedback from students has a positive impact on response quality and can boost students' enjoyment levels. Furthermore, gender differences and student age have been identified as important factors that can influence the depth of the conversation with the CA.

7. **Tsivitanidou O., Ioannou A.** (2021) Envisioned Pedagogical Uses of Chatbots in Higher Education and Perceived Benefits and Challenges. In: Zaphiris P., Ioannou A. (eds) Learning and Collaboration Technologies: Games and Virtual Environments for Learning. HCII 2021. Lecture Notes in Computer Science, vol 12785. Springer, Cham. https://doi.org/10.1007/978-3-030-77943-6_15 https://link.springer.com/chapter/10.1007%2F978-3-030-77943-6_15

Abstract. The widespread use of chatbots is a reality and their application in higher education is promising. Understanding higher education users' expectations for the use of chatbots in education is important for the design and development of new solutions. The present investigation documents how higher education users envision the pedagogical uses of chatbots in higher education, and how experts in the domain of education chatbots perceive the potential benefits and challenges related to the use of chatbots in education. A qualitative inquiry was undertaken based on 22 semi-structured interviews with higher-education students and instructors, and experts from the fields of Artificial Intelligence and educational chatbots. Based on our findings, the envisioned pedagogical uses of chatbots can be categorized in terms of chronological integration into the learning process: prospective, on-going, and retrospective. Under each one of those higher-order categories, specific learning domains can be supported (i.e., cognitive, affective), besides administrative tasks. Benefits and challenges foreseen in the use of pedagogical chatbots are presented and discussed. The findings of this study highlight the manner in which higher-education users envision the use of chatbots in education, with potential

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

implications on the creation of specific pedagogical scenarios, accounting also for the learning context, chatbot technology, and pedagogies that are deemed appropriate in each scenario.

8. **Frangoudes F., Hadjiaros M., Schiza E.C., Matsangidou M., Tsivitanidou O., Neokleous K.** (2021) An Overview of the Use of Chatbots in Medical and Healthcare Education. In: Zaphiris P., Ioannou A. (eds) Learning and Collaboration Technologies: Games and Virtual Environments for Learning. HCII 2021. Lecture Notes in Computer Science, vol 12785. Springer, Cham. https://doi.org/10.1007/978-3-030-77943-6_11 https://link.springer.com/chapter/10.1007%2F978-3-030-77943-6_11

   Abstract. Chatbots are becoming a trend in many fields such as medical, service industry and more recently in education. Especially in healthcare education, there is a growing interest in integrating chatbots in the learning and teaching processes mostly because of their portability and affordance. In this paper, we seek to explore the primary uses of chatbots in medical education, as well as how they are developed. We elaborate on current chatbot applications and research enacted in the domains of medical and healthcare education, We focus in the areas of virtual patients in medical education, patients' education related to healthcare matters but also chatbots as course assistance in for enhancing healthcare professionals' curricula. Additionally, we examine the metrics that have been used to evaluate these chatbots, which include subjective ones like the usability and acceptability by the users, and objectives ones, like their accuracy and users' skills evaluation. Overall, even though chatbots offer a flexible solution and a vast possibility to improve healthcare education, our literature review suggests that their efficacy has not been thoroughly tested. Also, limited examples of chatbots in European Healthcare curricula have been found. These call of the need for further research towards this direction.

9. **Antonio Miguel Mora, Alberto Guillén, Francisco Barranco, Pedro A. Castillo, Juan Julián Merelo Guervós** (2021) Studying How to Apply Chatbots Technology in Higher-Education: First Results and Future Strategies. In: Zaphiris P., Ioannou A. (eds) Learning and Collaboration Technologies: Games and Virtual Environments for Learning. HCII 2021. Lecture Notes in Computer Science, vol 12785. Springer, Cham. https://doi.org/10.1007/978-3-030-77943-6_12 https://link.springer.com/chapter/10.1007%2F978-3-030-77943-6_12

   Abstract. This paper tries to find the best condition to use chatbots (conversational agents) in higher-education studies after pilots carried out at the University of Granada (Spain). Our aim, along with the rest of partners in EDUBOTS -an Erasmus + European Project which counts with two pedagogical chatbots-, is to improve students' engagement in class, as well as reducing the existing gap between them and their educators. In this paper we present the results of a previous survey carried out among the students with the intention of laying out a plan of possible effective applications of this technology in the classroom in the near future, if possible during the next project pilot. The survey helps us

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

confirm the reasons for the learning outcomes in the carried out pilots, as well as identify the targets for future application of chatbot technology.

10. **Dijana Oreški, Dijana Plantak Vukovac, Goran Hajdin** (2022). Assessment and comparison of meta-features for educational chatbots data and survey data, *WSEAS Transactions on Computers* (accepted for publishing)

    Abstract. Usage of chatbot platforms is acquiring great attention at all levels of education. Human-chatbot interactions generate huge amounts of data which are a valuable source of information, when properly analyzed by means of data and text mining. One of the most challenging tasks in the mining process is the selection of the appropriate algorithm for a data set at hand. This is a complex task and depends on characteristics of the dataset used in the analysis. Those characteristics are formalized through meta-features. In this paper, we identified meta-features of chatbot and survey data. As a case study, we evaluated two data sets and identified their general meta-features along with discussion. This is, to the best of our knowledge, the first examination of meta-features for chatbot interactions data and their comparison with survey data.

11. **N Abbas, J Whitfield, E Atwell, H Bowman, T Pickard A Walker.** (2022) Online Chat and Chatbots to Enhance Mature Student Engagement in Higher Education. Submitted to *International Journal of Lifelong Education*, https://eprints.whiterose.ac.uk/

    Abstract. Mature students transitioning into their first year of higher education face many difficulties that affect their motivation, participation and success. Their feelings of being disconnected from their peers and from their institutions, in addition to the challenges they face while managing their finances and family responsibilities, are among the key barriers to the successful completion of their courses. Encouraging online student engagement among mature students to establish social connection with their instructors and peers can reduce their isolation and enhance their 'sense of belonging'. During the academic year 2020/2021, the Lifelong Learning Centre (LLC) at the University of Leeds in the UK, decided to pilot an online chat platform 'Differ' including a chatbot 'Bo' after it had seen a decline in the use of programme Facebook groups. To promote student engagement and monitor the Differ online communities, sixteen digital student mentors were recruited and trained. During the pilot, feedback was collected from students and student mentors. A mixed-methods approach was adopted to boost robustness through triangulation. Several forms of data collection methods were used: a mentimeter, an online survey, three focus group sessions and a semi-structured interview. This study sheds light on the different aspects of creating student-led online communities and provides recommendations on how to improve the uptake of students in the future.

12. **J.J Merelo, P. A. Castillo, A. M. Mora, F. Barranco, N Abbas, A. Guillen, O. Tsivitanidou** (2022) Exploring the Role of Chatbots and Messaging Applications in Higher Education: A Educator's Perspective. Accepted to be presented at HCII 2022, https://2022.hci.international/

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Abstract. Introducing new technologies such as messaging platforms, and the chatbots attached to them, in higher education, is rapidly growing. This introduction entails a careful consideration of the potential opportunities and/or challenges of adopting these tools. Hence, a thorough examination of the educators' experiences in this discipline can shed light on the effective ways of enhancing students' learning and boosting their progress. In this contribution, we have surveyed the opinions of tertiary education educators, who are based in Spain (mainly) and Spanish-speaking countries. Specifically, we aimed to collect educators' feed-back about their opinions regarding the introduction of the messaging platforms and chat-bots in their classes, understand their needs and to gather information about the various educational use cases where these tools are valuable. In addition, an analysis of how and when educators' opinions towards the use of these tools can vary across gender, experience, and their discipline of specialization is presented. The key findings of this study highlight the factors that can contribute to the advancement of the adoption of messaging platforms and chatbots in higher education institutions to achieve the desired learning outcomes.

13. **J.J Merelo, P. A. Castillo, A. M. Mora, F. Barranco, N Abbas, A. Guillen, O. Tsivitanidou** (2022) Chatbots and messaging platforms in the classroom: an analysis from the educator's perspective. Submitted to Journal International Journal of Educational Technology in Higher Education.

    Abstract. Final abstract Non-Available

14. **A. M. Mora, F. Barranco, A. Guillén, J.J Merelo, P. A. Castillo** (2022) Estudio sobre el uso de chatbots en docencia dentro de titulaciones de la ETSIIT de Universidad de Granada (*Study on the use of chatbots for teaching in different degrees at ETSIIT at the University of Granada*). Accepted for publication in Revista de Enseñanza y Aprendizaje de Ingeniería de Computadores (*Journal of Teaching and Learning Computer Engineering*).

    Abstract. Este trabajo presenta las experiencias de cinco profesores de la Escuela Técnica Superior de Ingenierías Informática y Telecomunicaciones (ETSIIT) de la Universidad de Granada, dentro del proyecto Erasmus+ EDUBOTS durante 2 cursos académicos. En él se pretendía aplicar la tecnología de los chatbots (o agentes conversacionales), como herramienta de apoyo (y motivación) para estudiantes y profesores de diferentes grados y másteres dentro de la escuela. Durante dicha experiencia se utilizaron dos aplicaciones específicas que incorporan chatbots (Differ y Hubert), proporcionadas por empresas participantes en el proyecto, y además se hizo uso de una aplicación de mensajería instantánea, como es Telegram, junto con varios de los bots que permite desplegar y utilizar. Los resultados de los pilotos realizados dejan entrever que los estudiantes no se sintieron motivados para el uso activo de dichas herramientas, salvo Telegram, en las últimas experiencias llevadas a cabo. Los profesores por su parte, también opinan que este tipo de propuestas no resultó efectiva, pero plantearon distintas posibilidades de aplicación que podrían ser más exitosas.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

This paper presents the experiences of five professors of the School of Computer and Telecommunication Engineering (ETSIIT) of the University of Granada, within the Erasmus+ EDUBOTS project during two academic years. The aim of the project was to apply chatbots *(conversational agents)* technology as a support (and motivation) tool for students and *educator*s of different degrees and masters within the school. During this experience, two specific applications that incorporate chatbots *(Differ and Hubert)* were used. They were provided by companies participating in the project. In addition, Telegram, an instant messaging application, was also used, along with several of the bots that can be deployed and used inside it. The results of the pilots carried out suggest that students did not feel motivated to actively use these tools, except Telegram, in the last experiences conducted. *Educator*s also felt that these *types* of *approaches were* not effective, so they proposed different application possibilities that could be more successful.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

Conclusion

In the EDUBOTS project new technologies are used to enhance student experience and to facilitate interactions between students and educators. Two applications were predominantly used as a part of the project: Differ and Hubert, but also other applications like Telegram were explored in terms of chat and chatbot functionalities, and their role in teaching and learning process. Those applications were used in higher educational courses at the universities in Croatia, Cyprus, England and Spain through four pilot rounds, from January 2020 till December 2021.

This document presents best cases and practices in using chat applications and chatbots, one case per partner institution in each pilot. Best cases are presented in a narrative form giving qualitative insight of onboarding educators and students, pedagogical approaches and ways of implementation of the applications.

By using chatbot application Differ, educators created informal course communities for communication with students. It helped students to connect with other peers but also educators. It was a great opportunity to have a more open conversation upon class matters, make projects or study groups and discuss other topics related to personal interests (hobbies, music etc.). **Onboarding** was the task of every course educator, and invitation messages for students were sent through the learning management system (Moodle or Blackboard), email or in lectures. In total, Differ was used in almost 50 courses across four pilots.

The role of chatbot application Differ and chatbot BO in the **learning process** were examined in all four Pilots from two standing points: (I) to increase student engagement in course and (II) to help students in bonding with peers and educators to develop social connections and sense of inclusion to the faculty. Through pilots and role of application and later, chatbot BO, students perceived Differ as a safe environment in which they can connect better with course educators and peers. Students evaluate Differ very positively for communication with educators during the COVID-19 pandemics, so it proved an appropriate tool for fostering academic activities. Due to learning activities, students recognized a Differ supporting role for collaboration with colleagues, completing and participating in course activities, but its direct impact for achieving better grades in the course was not strongly reported in all four Pilots. Differ's role in pandemics in the mostly online teaching environment was accomplished successfully as one of its **main pillars** - to **provide safe online space** for student's formal and informal communication and increased social inclusion of each individual in the community.

As a part of the pilots we systematically performed analysis of the educators' and students' use of Differ and Hubert. Main goal of the results presented in sections 5.3.1/2, 6.3.1/2 and 7.3.1/2 was to explore how students and educators use Differ and Hubert and how the apps helped them in the process of learning and teaching. Conducted surveys had a rather constant low number of participants which mostly consisted of participants from the University of Zagreb, Faculty of Organisation and Informatics. With the amount of changes which were implemented in the

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

universities, especially due to the COVID pandemic, as well as follow-up surveys, it was hard to motivate them to participate in the questionnaires.

The results are mixed, depending on the participants. Due to the low number of respondents we can't make any firm statistical conclusions. Most of the participants from project partner countries in all four pilots were students from one university which is also one of **the limitations of project** results and overall research contributions. Based on the results, most users used the app on a weekly basis and they dominantly used desktop and mobile versions. Rather constant positive results were related to the perception that educators would like to use Differ in other courses, they felt free to express their thoughts, they agreed that students didn't use inappropriate language or communication within the app, and they would recommend the app to their colleagues. Other results which are focused on the teaching process have a wider distribution of answers, adding to the point that there is no single "golden" method, approach nor app which all educators would use the same way.

Related to the Hubert results, educators constantly agree that the setup process, as well as download and access to the evaluation results were easy. Educators have a more neutral opinion towards the thematic analysis of the results which is presented in the Hubert, since the analysis can be properly conducted only with the results which are entered in English language.

For the BO functionalities, educators liked the most the creation of community for students. Other two functionalities which were positively accepted were asking questions to the community and splitting the students into groups. Functionalities which had mixed results were related to the introduction activities (ice-breaking) and frequently asked questions (FAQ).

In the scope of Pilots round 2 and 4, interviews with students were conducted. Focus of the interviews was to examine experience with chatbots and expectations related to chatbot usage. Interviewees stated that their experiences were mainly positive, but they weren't completely satisfied and suggested improvements in terms of predefined answers, speech recognition and bot's personality. Students mostly tend to use chatbots in their lives in the future.

Interviews with educators were also conducted in Pilots 2 and 4. Functionalities and experience of usage of Differ and chat bot BO in various educational case scenarios were examined. Analysis of interviews resulted with valuable feedback in the terms of technical, usability and user experience improvements that will be taken into consideration before releasing new versions and functionalities of Differ, BO and Hubert. On the one hand, educators emphasised advantages such as clear and intuitive interface, easiness to use and understand. On the other hand, educators recognize weaknesses which require better developed functionalities and should be taken into consideration.

Across all four pilots, we explored usability and user experience of both Differ, BO and Hubert. In Pilot 1 and 2 focus was put on the usability of Differ evaluated with SUS questionnaire and user experience evaluated with User Experience Questionnaire (UEQ) that measures both usability and hedonic quality aspects. According to the mean SUS score, in Pilot 1 Differ was evaluated as a slightly above average application, but it was evaluated a bit lower in Pilot 2 than in Pilot 1. On the UEQ scales, attractiveness and pragmatic quality were scored positively. However,

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

hedonic quality (stimulation and novelty) was scored below average in Pilot 1, but it gained neutral to positive marks in Pilot 2, probably because of the introduction of BO. So, in Pilots 3 and 4 we evaluated BO with SUS, UEQ and Chatbot Usability Questionnaire (CUQ), the latter one comparable to SUS with addition of statements related to chatbots. BO did not gain better scores than Differ in Pilot 1 and 2 with SUS and UEQ, but scored better than average with CUQ for its welcoming during initial setup, ease of use, usefulness, appropriateness, informativeness and friendliness in Pilot 3 and 4. According to those results there is a lot of room for improvement in both user interface and functionality of Differ and BO, in order to become attractive enough for students who are accustomed to other general chat applications.

Usability of the chatbot Hubert was evaluated with two short questions from the UMUX-LITE questionnaire, which were integrated at the end of the conversation with Hubert. Unfortunately, those questions were implemented with the answers on the reverse scale (1 for the highest score, and 7 for the lowest score), so this might have confused some students and the results might not be valid. However, Hubert was evaluated with CUQ in Pilot 3 and 4, giving better than average score for its welcoming during initial setup, ease of use and ease of navigation. Finally, Hubert was perceived as a simple chatbot with a defined purpose. With the UEQ, Hubert gained all positive evaluations except for stimulation, so new improvements might go into implementing more inspirational or intriguing questions and answers.

With the project objectives in mind, ten research papers were published and presented at the scientific conferences. In addition, four papers are submitted to the journals, with one already accepted for publishing. All those papers present part of the pilot results and identify directions for further research and improvements.

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN

References

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, *24*(6), 574–594. https://doi.org/10.1080/10447310802205776

Díaz-Oreiro, López, Quesada, & Guerrero. (2019). Standardized Questionnaires for User Experience Evaluation: A Systematic Literature Review. *Proceedings*, *31*(1), 14. https://doi.org/10.3390/proceedings2019031014

Hassenzahl, M., & Tractinsky, N. (2006). User experience—A research agenda. *Behaviour & Information Technology*, *25*(2), 91–97. https://doi.org/10.1080/01449290500330331

Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V., & Mctear, M. (2019). Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? *Proceedings of the 31st European Conference on Cognitive Ergonomics*, 207–214. https://doi.org/10.1145/3335082.3335094

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013). UMUX-LITE: When there's no time for the SUS. P*roceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2099–2102*. https://doi.org/10.1145/2470654.2481287

Martín, J., Muñoz-Romero, C., & Ábalos, N. (n.d.). Chatbottest—The free guide for you to understand what is your chatbot doing wrong. Retrieved 6 January 2022, from https://chatbottest.com/

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence*, *4*(4), 40. https://doi.org/10.9781/ijimai.2017.445

User Experience Questionnaire (UEQ). (n.d.). Retrieved 12 January 2021, from https://www.ueq-online.org/

Co-funded by the Erasmus+ Programme of the European Union

THIS PROJECT HAS BEEN FUNDED WITH SUPPORT FROM THE EUROPEAN COMMISSION UNDER THE ERASMUS+ PROGRAMME. THIS PUBLICATION [COMMUNICATION] REFLECTS THE VIEWS ONLY OF THE AUTHOR, AND THE COMMISSION CANNOT BE HELD RESPONSIBLE FOR ANY USE WHICH MAY BE MADE OF THE INFORMATION CONTAINED THEREIN