DELIVERABLE D5.5: PILOTING AND EVALUATION REPORT 2

Document Version:	V3– March 6, 2016
Document Status:	Version 2
Document Type:	Report
Diss. Level:	Public
Lead Partner:	SCIO
Delivery Date:	M24 (February 2016)
Authors:	Andrej Novik (SCIO) Lenka Firtova (SCIO) Mustafa Ali Türker (SEBIT) Michael Kickmeier-Rust (TU Graz) Michael Bedek (TU Graz) Matthew Johnson (UoB) Blandine Ginon (UoB) Susan Bull (UoB)

TABLE OF CONTENTS

1. INTRODUCTION AND SCOPE

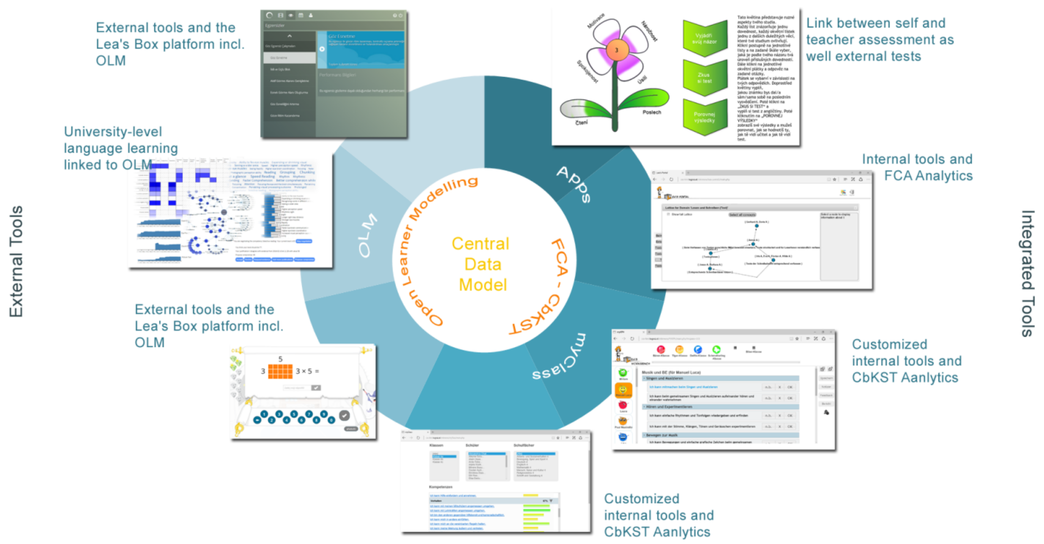
Deliverable 5.5 presents the second stage of piloting of the LEA's Box tools. Piloting is one of the most important parts of the LEA' Box project: after all, no matter how successful research results may seem, it is always crucial to try out the tools and applications developed "in a lab" in the real world.

This deliverable is a continuation of D5.3, which was very practically oriented, unlike deliverables D5.1, which focused on the piloting methodology, D5.2, which proposed activities for pilot studies, and D5.4, which described different scenarios and classroom issues. Deliverable D5.3 described several pilot activities and tools, such as the myClass tool or the mind-mapping tool. In the meantime, the tools developed within the project were substantially improved, new tools were built on the existing ones and new use-cases were proposed and tested. These activities were preceded by a thorough evaluation of the feedback generated during the first phase of piloting.

Therefore, this deliverable (D5.5) describes new extensions of the LEA's Box tools and new use-cases and explains what changes have been made and why, i.e. what steps we took in order to reflect teachers' and students' needs in the best possible way. Namely, the following is going to be presented:

- pilot activities done by Scio in the Czech Republic
- pilot activities done by Sebit in Turkey
- a use-case proposed by TU Graz
- a study conducted by the University of Birmingham

The following image illustrate the dimensions of the use case studies, ranging from internal applications to integrating external tools, from activity tracking to visualizatiosn and Open Learner Modelling.



2. PILOTING AND EVALUATION IN THE CZECH REPUBLIC (SCIO)

2.1 SUMMARY OF TEACHERS' AND PUPILS' NEEDS

Following pilot studies described in Deliverable D5.4, we decided to reanalyze teachers' and pupils' feedback regarding their needs, requirements and their satisfaction with the tools that had been developed earlier before proceeding to the second stage of piloting in order to be able to shape our course accordingly. One source of such information were focus groups and an online survey filled in by teachers from 88 different schools in June 2015. Another source of information were structured interviews conducted from October to December 2015 in 23 different schools:

- Základní škola Chomutov
- ZŠ Palachova, Ústí nad Labem
- ZŠ Chotěšov
- ZŠ Martina Luthera, Plzeň
- ZŠ Čechtice, okr. Benešov
- Základní škola a mateřská škola Čkyně
- Základní škola Třeboň, Sokolská 296
- ZŠ Londýnská
- Přírodní škola, Strossmayerovo náměstí 990/4, Praha 7
- Základní škola, Praha 3, nám. Jiřího z Poděbrad 7,8/1685
- Základní škola, Praha 2, Sázavská 5
- ZS Da Vinci
- ZŠ Nebušice
- G MensaŠpanielova 1111/19, Praha 6
- ZŠ Mills Čelákovice
- Masarykova základní škola a Mateřská škola Debř
- ZŠ Čeperka
- ZŠ Malíka Chrudim
- ZŠ T.G. Masaryka, Náchod
- Základní škola a Mateřská škola Stárkov
- ZŠ Bruntál, Jesenická 10
- ZŠ Mládežnická, Havířov
- ZŠ a MŠ Karlovarský region, Sadov

The aim of these interviews was twofold: first, as mentioned above, to get more information about teachers' and pupils' real needs and requirements, and second to support the dissemination of the LEA's Box project, both being successfully accomplished.

Let us now summarize the main findings which formed the basis of our further work.

Finding I.: PCs and laptops are used much more frequently in the classrooms than tablets

This can be seen from Figure 1, which was already included and thoroughly commented in D5.3. Only a few teachers reported they use tablets in the classrooms on a daily basis, while as many as 50 % of them said they use PCs or laptops several times a week or even daily. This implies teachers would probably currently appreciate tools suitable for PCs or laptops rather than tablet applications.

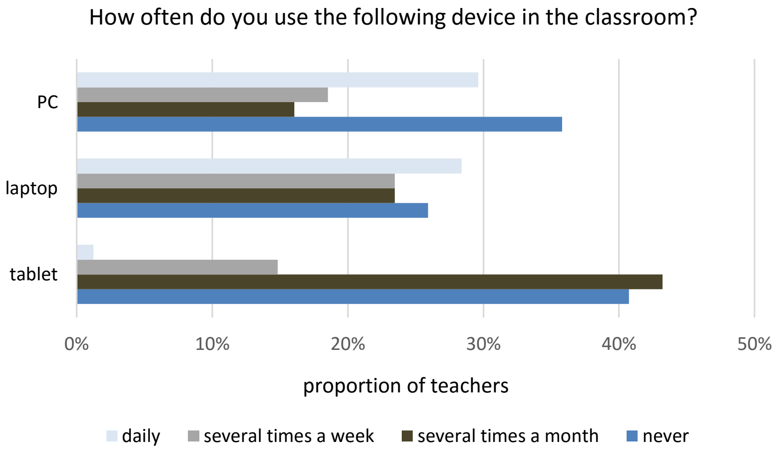


Figure 1. Frequency of use of different types of devices

Finding II.: Above all, teachers miss suitable tools for self-assessment

While teachers do recognize the importance of project-based learning and of tracking students' progress on a continuous basis, they typically already have at least some basic tools do so, although there is, of course, room for improvement. However, one thing was mentioned repeatedly during the interviews, and that is the need for a tool which would allow students to assess themselves and to compare self-assessment with their teacher's opinion and with the results provided by an external evaluation tool (such as a standardized test). While generally, teachers did appreciate the usefulness of the mind-mapping tool, the MyClass tool and the concept of digital rubrics of learning evidence, which were all described in detail in D5.3, they said that currently, they would prefer to get a tool which would put more focus on self-assessment.

Based on these findings, we decided to develop an extension of an existing tool, SCATE (Scio Computer Adaptive Test), in order to combine self-assessment and teachers' assessment with information provided by this test, as we concluded that this is what teachers would currently appreciate the most.

2.2 Pilot studies

2.2.1 Flower Tool

The Flower Tool is a direct product of our discussions with teachers, especially from ZŠ Londýnská, who described what such a tool should ideally look like. Although it works on tablets, it is mainly suitable for PCs/laptops. As mentioned above, the main aim of the Flower Tool is to allow students to compare their own opinion of their abilities with their teacher's opinion and with the results of an external test. For piloting, the skills we chose to measure were reading and listening skills in English, as we already possessed a suitable test evaluating these skills. In the future, it will be possible to link any suitable test to the application.

The application works as follows. There is a flower with leaves representing different skills (in this case, reading and listening) and with petals representing other important aspects that may influence learning progress (in piloting, we used motivation, effort, satisfaction and load, but in the future it will be possible for teachers to modify these as they deem fit). It is a widely recognized fact that it is necessary to pay attention not only to the learning progress itself, but also to what may influence this progress, such as load or motivation (Sweller, 1994, Ames, 1992). Teachers themselves repeatedly complained they do not have tools which would allow them to evaluate not only their students' learning progress per se, but also factors which may influence it. That is why we consider the petals especially important.

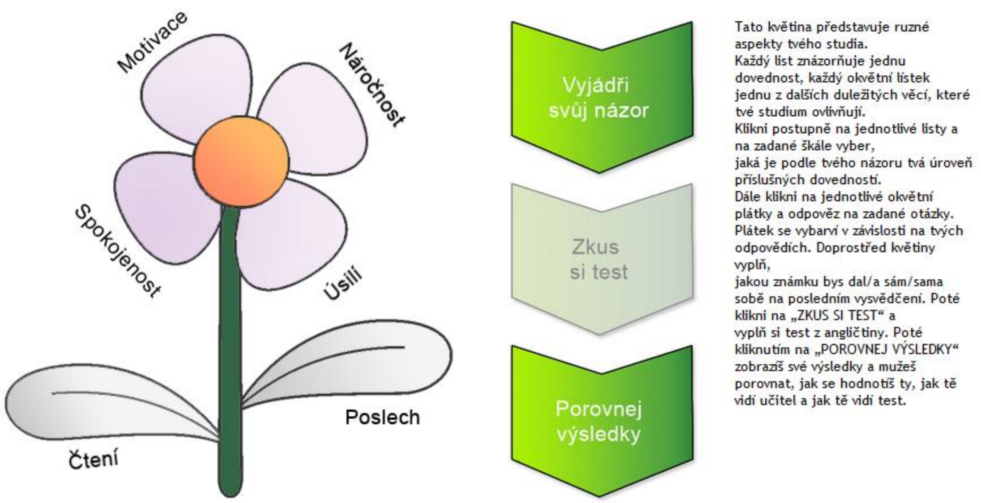


Figure 2. Flower Tool

First, students click on the leaf representing reading ("Čteni" in the picture above) and say what they think their level of reading skills is by choosing a corresponding description of what particular skills they should display if being on such a level (e.g. "I can understand short texts on common everyday topics"). The descriptions are based on the Common European Framework, i.e. each description corresponds to one level from A0 (we introduced this level as a level describing complete beginners) to C1. Based on their answer, the leaf is filled with colour, where A2 means that 40 % of the leaf should be coloured in, B1 means that 60 % of the leaf should be coloured in etc. The same is then done for listening ("Poslech").

Next, students click on the petals. When clicking on a petal, a short questionnaire appears which helps students determine the level of their motivation, effort, satisfaction and load. An example of the questionnaire for motivation is shown below. Based on their answers, a corresponding proportion of the petal is filled in with colour (e.g. if a student chooses 6 in each of the questions, the whole petal turns violet). Students, of course, are allowed to modify their answers.

MOTIVATION	1 = strongly disagree			5 = strongly agree		
1. In my opinion, English is the most important subject at school.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
2. I enjoy learning English.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
3. Out of all the subjects, I find it most important to improve my English.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6

Last but not least, students enter the grade they believe they deserve into the centre of the bloom (in the Czech educational system, it's typically a number from 1 to 5).

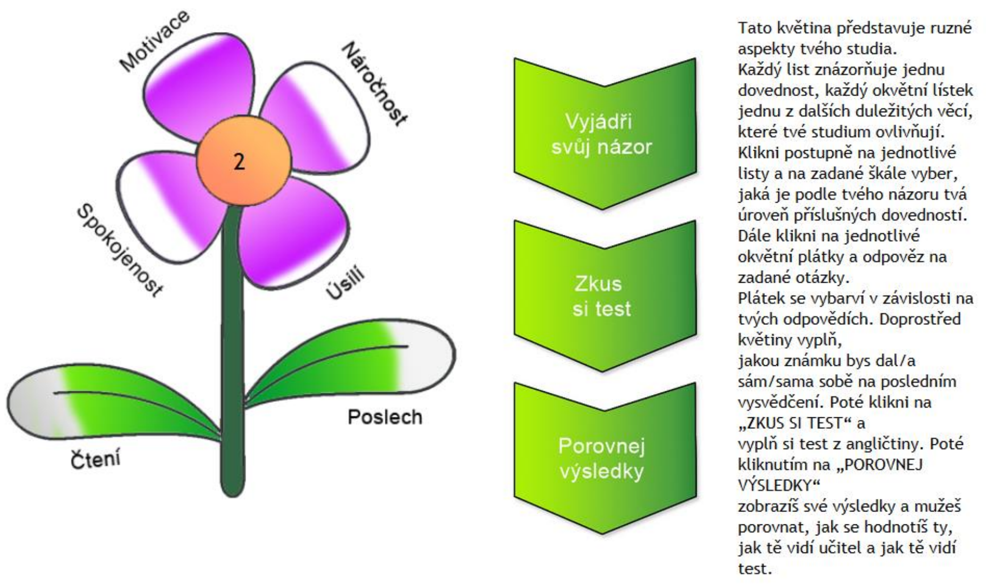


Figure 3. Example of a flower which has been filled in

Then, students take SCATE, a test developed by Scio, which puts students into one of the categories defined by CEFR (A1, A2, B1, B2, C1) or the A0 category (complete beginners) in both reading and listening. The test is adaptive and includes grammar tasks, conversation tasks, vocabulary tasks and gist/detail tasks, so the level of each student is determined with sufficient precision.

Meanwhile, teachers assess listening, reading, motivation, load, effort and satisfaction of each of their students themselves (teachers see the same description of the reading and listening skills and the same questionnaires regarding motivation etc., but without the flower).

After all of this has been done, students can compare all the three sources of information about their performance, and so can their teachers. The results are displayed as follows. The first four criteria are motivation, effort, satisfaction and load, the green bar indicating what a student thinks and the blue line indicating what a teacher thinks (for these four criteria, test results are not available). The next two criteria are reading ("Čtení") and listening ("Poslech"), the green dots, again, indicating the opinion of a student, the yellow dots the results the student got in the test and the blue dots reflect the opinion of the student's teacher. The student can also see these results in the form of a radar graph and a column graph.

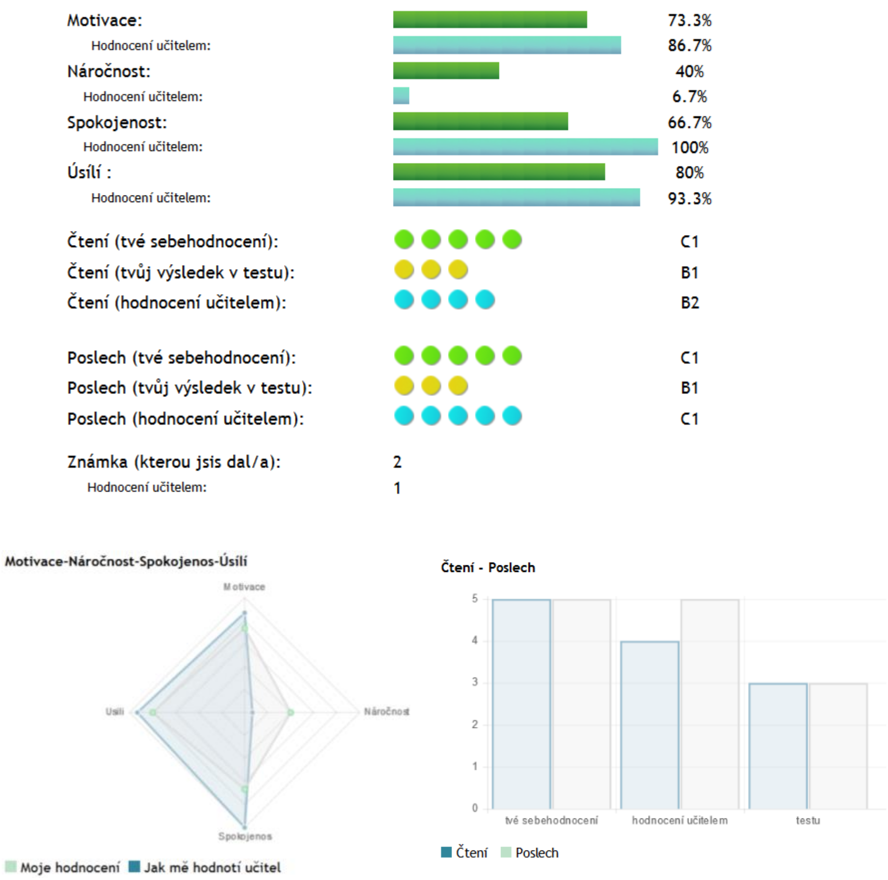


Figure 4. Example of results shown to the student and graphs generated by the Flower Tool

Teachers can display results for each student separately in the form of a table and a graph. An example, this time in English (as the teachers are fluent enough to work with the English version of the application) are shown below. Teachers can also display a summary for the whole class, both as a table, and as a graph for a particular skill (e.g. reading).

Summary I (in percentages):

	Motivation	Load	Satisfaction	Effort
Student	73.3	40	66.7	80
Teacher	86.7	6.7	100	93.3

Summary II:

	READING	LISTENING	GRADE
Student	C1	C1	2
Teacher	B2	C1	1
Test	B1	B1	

Figure 5. Example of results shown by the Flower Tool to the teacher

2.2.2 PILOTING

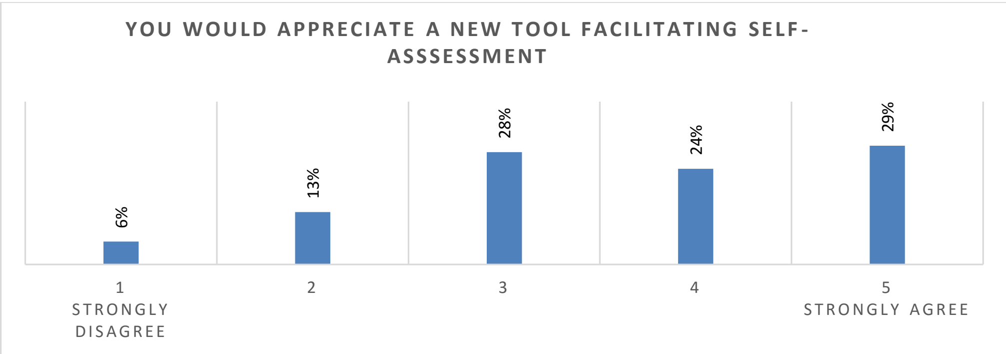
Altogether, 559 pupils aged 10 to 15 from four different schools participated in piloting, which took place in February 2016. 326 of them went through the whole application including a final questionnaire designed to get their feedback, and some are going through the piloting process these days, so the final numbers may be even higher. The results of the piloting (taking into account these 326 pupils) are described in the next chapter.

2.3 EVALUATION AND VALIDATION

The aim of this chapter is to evaluate the feedback we gathered when piloting the Flower Tool. Our first source of information were the students themselves who were presented with a thorough questionnaire regarding mainly the assessment practices in their school and the Flower Tool itself.

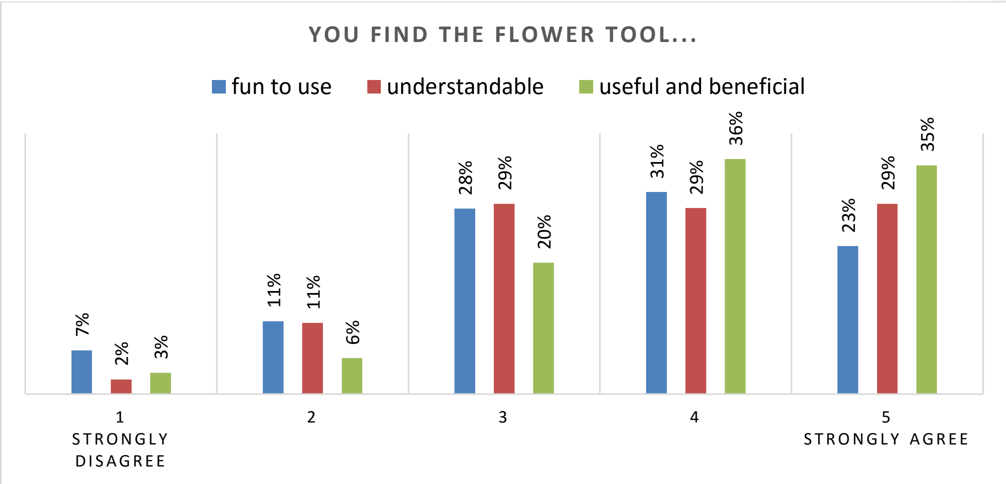
The first set of questions concerned the use of ICT in the classroom and the results support our previous assumptions that PCs and laptops are used much more frequently in the classrooms than tablets: only 29 % of pupils said they never use a PC or a laptop in the classroom, but as many as 74 % of them said they never use a tablet.

The second set of questions focused on students' needs with respect to assessment, and, once again, the findings support our previous claim that what is needed most in today's schools are suitable tools for self-assessment. The first graph below shows that, unfortunately, only 26 % of students think self-assessment gets enough space in their school and, similarly, only 24 % of them said they can compare self-assessment with their teacher's assessment frequently enough. The majority of students has no strong opinion on this issue and about a third of them said they disagree with the statements which are included in the following two graphs. Last but not least, it can be seen from the third graph that most students indicated they wanted to engage with self-assessment with the support of technology, in addition to the facilities currently available to them.



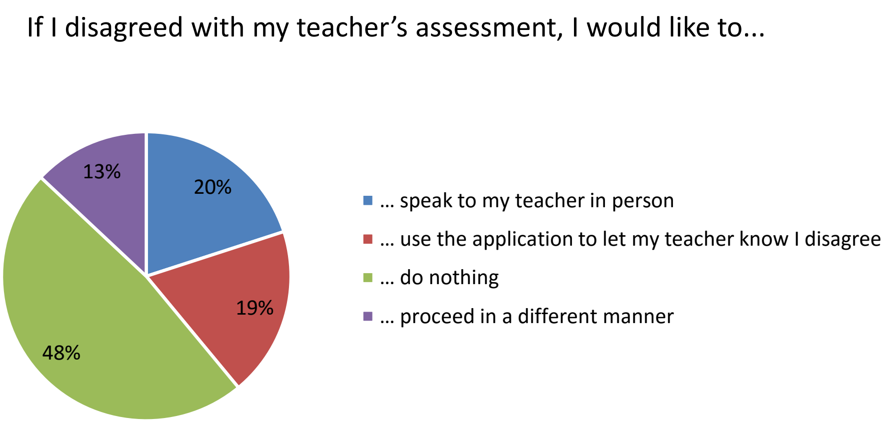
Graphs 1. to 3: Pupils' opinion on self-assessment related issues

The third part of the questionnaire concerned the Flower Tool itself. In this respect, the feedback was generally very positive. More than half of the students said they enjoyed working with the tool, that they understood what they were supposed to do, and, most importantly, that they recognized the usefulness of this tool. Compared to the tools we piloted last time, the myClass tool and the mind-mapping too, this feedback is a significant improvement. Last time, it was mostly the teachers, some from the same schools and some from different schools, who evaluated the tools, and the prevailing opinion was that the tools were not as user-friendly as one might expect and that it was it would be great if they involved self-assessment (see D5.3 for details). From the feedback to the Flower Tool it is clear that a thorough analysis of the feedback received after the first stage of piloting, which was conducted prior to the second stage of piloting, proved very effective.



Graph 4: Pupils' feedback regarding the Flower Tool

Next, we wanted to get feedback regarding the need to introduce tools that facilitate the negotiation between a student and the application. Therefore, we asked students what they would prefer to do if they disagreed with their teacher's assessment of their work displayed by the application. 20 % of students said they would speak to their teacher in person, 19 % said they would like to use the application in order to initiate negotiation (i.e. they would like to let their teacher know they disagree with his or her assessment of their work using the application), 48 % of them said they wouldn't do anything and 13 % said they would proceed in a different manner, as shown in the graph below.



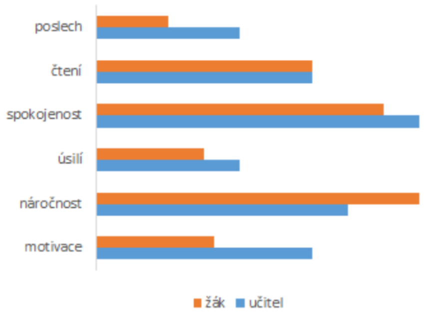
Graph 5: Pupils' preferred course of action in case of assessment discrepancy

This shows that only a handful of students would actually appreciate using the application in order to get their assessment modified, which led us to the conclusion that currently, we do not need to implement a negotiation tool into the Flower Tool. In addition, we believe it is wiser to present students with one new concept first, i.e. self-assessment and its comparison with other sources of assessment, in order to give them time to get used to it, before introducing new functionality.

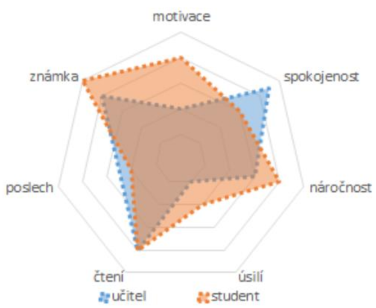
Last but not least, students were presented with a set of graphs which depicted the comparison of self-assessment and teacher's assessment. They were asked how understandable the graphs were and which one they preferred. Graph 2 and Graph 4 are quite similar – both offering quite a straightforward comparison of different points of view. Graph 1 might be less understandable for children, as the values are displayed as a continuous variable, which doesn't correspond to the reality. Graph 3 is a very useful way to represent information, as it, when read properly, easily shows two different kinds of information at the same time – when we look uniquely at the orange (student's) radius, we can see how the performance in different domains (motivation, reading etc.) differs in the eyes of this student (what he or she thinks they are best at etc). and the same holds true for the blue radius representing the teacher's point of view. But also, comparing the two radii allows us to directly compare these different points of view at a certain domain. Unfortunately, students are not used to working with this type of visualization much and may have a hard time interpreting what it represents.

Out of the four graphs offered, students preferred Graph 4, with as many as 47 % of them saying they liked it more than the other graphs. Graph 4 was followed by Graph 2, which was chosen as the best by 29 % of the students, see below for full results. These findings are consistent with earlier studies (see Coury, Boulette & Smith1989).

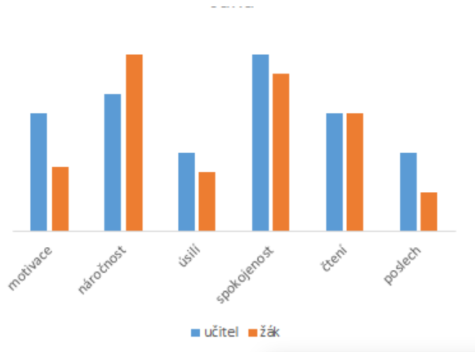
Graph 1



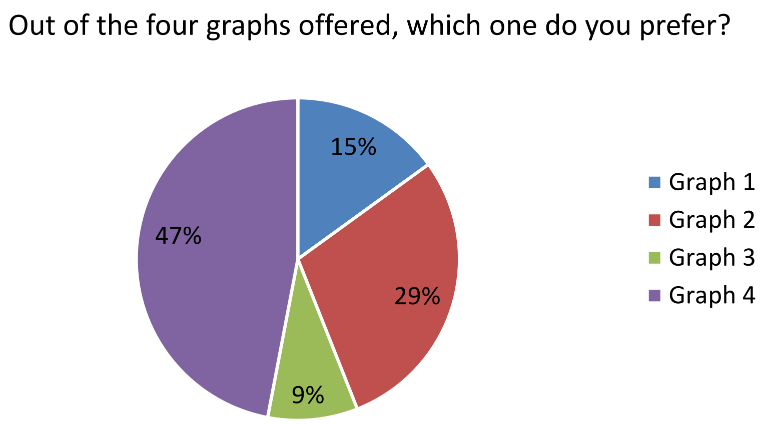
Graph 2



Graph 3



Graph 4



Graph 6: Pupils opinion on different types of visualization

Our second source of feedback are teachers. As we prefer direct interviews, focus groups and skype conferences with the teachers to one-way questionnaires, we are going to get most of this feedback in person or over the phone / Skype. The aim of this interaction will be to get thorough responses to the questions in the questionnaire together with a detailed commentary and to discuss other possible problems, issues and suggestions for improvement, as we know that this is crucial before developing any new application or improving an existing one. However, so far, no teacher has reported any significant problem or issue regarding the Flower Tool. These interviews and focus groups will be complemented by workshops to support dissemination and exploitation activities. The first workshop took place on January 16 in ZŠ DaVinci Dolní Břežany. Here we discussed, among other things, the functionality of the Flower Tool, and we looked for ways to improve it in order to make it as beneficial for schools as possible.

2.4 Summary of the benefits of Flower Tool and further development

To sum up, between the first and the second stage of piloting, we took time to reanalyse teachers' and pupils' needs and requirements and concluded we should come up with a tool facilitating self-assessment and comparison of different sources of assessment. We developed this tool in cooperation with TU Graz, after discussions with teachers about what the tool should ideally look like and what functionality it should have. As we didn't want to develop a completely new tool from a scratch, we used an existing test developed by Scio and the Flower Tool became an extension of it.

In general, the main benefit of the Flower Tool is that it meets the requirements of the teachers and students in several ways. Firstly, it promotes the concept of self-assessment: once self-assessment becomes a regular part of the learning process thanks to the Flower Tool, it is likely to gain importance and become part of other activities which teachers design. Thus, we strongly believe the Flower Tool is going to promote the use of self-assessment in the classrooms in general.

Secondly, it combines different sources of assessment, in which it is unique and different from other existing tools that may be used for self-assessment as well. This comparison allows teachers to identify possible sources of problems, both related to the subject itself (e.g. English) and to their students' attitudes and behaviour. For example, thanks to the Flower Tool, a teacher may find that a student keeps underestimating him/herself and take appropriate action. Without a suitable application, students who tend to underestimate themselves would not be very likely to come to their teachers themselves to look for a solution and teachers may be having a hard time looking for the sources of their unsatisfactory learning progress. Another positive aspect of the tool is that it also measures motivation, effort and other relevant factors, and, once again, compares students' opinions with their teachers' opinions. Thanks to this, a teacher may find that a student is struggling in a particular subject even though it may not be obvious in the classroom, or that they are not motivated enough. That's why the Flower Tool may shape students' progress not only in a particular subject, but it may also have a positive effect on their attitudes and behaviour, which may sometimes be even more important, as attitudes towards learning affect the learning process as a whole.

In the future, our main aim is to improve the tool based on the feedback we receive, namely:

- to make the tool more flexible, so that teachers can use it for different subjects, change the domains they want to test or change the design to something else than a flower (one of the most common suggestions by children was a unicorn);
- to modify the tool in such a way so that it would be suitable for formative assessment, i.e. to include functions which would track the progress of a particular student in time;
- to improve the visualizations so that they correspond to what students and teachers understand best.

After all, the Flower Tool is primarily made for teachers and students, and it's them who should decide what it will look like and what it should be able to do.

3. PILOTING AND EVALUATION IN TURKEY (SEBIT)

3.1 SUMMARY OF TEACHERS' AND PUPILS' NEEDS

The concrete use cases given in D2.2 Revised System Design Document, and the pilot studies on the first release of the system as described in Deliverable D5.4 point to **four** major findings which were taken to the piloting of the second release of the system.

Finding I: Competency-based performance analytics on personal basis helps students to remain on track by revealing the cause-effect relationships between which activities they did and what competency gains they achieved.

Learning is a constructive process. "Being on track" is a comprehensible concept as to whether this process is running fine or not. To be on track is, in general, relative to the course plan and the status of the classroom. To be advanced on the track, or retarded or being off-track are also easily comprehensible concepts. What's hard for students to comprehend is to "keep the track" based on competencies. As students are accustomed to being evaluated per study area, competency-based tracking is and it general "what a competency is" is an idea they need to cultivate.

Finding II: Teachers find real-time analytics disturbing since it prevents them to ponder longer term trajectories, but when multisource analytics is provided, such as attendance, and time domain views are presented, real-time analytics gain context.

To make any causal judgement based on analytics, time domain views are necessary. Most decisions about how to engage for class, what to assign, which activity to follow are all made based on some causal inferences. To make "data wise" inferences, the data needs to be presented with a time component, so that the effects of past decision can be evaluated into better inferences.

Finding III: As stakeholders, publishers and course developers are direct beneficiaries of analytics results, since it enables them to validate the competency structure, improve the course plan, as well as activity designs.

As a publisher and digital content developer, SEBIT can be consider a direct stakeholder for learning analytics tools. SEBIT content and course plans are structured based on competency maps, with a constructivist approach.

Finding IV: Evaluating evidences helps draw causal relations but to engage in negotiation there has to be a high enough discrepancy between which level the student feels at and which level the model displays.

Some students are interested in seeing the raw evidential data themselves rather than through the lens of analytics in order to make inferences. However, most of them don't engage in negotiation by themselves unless there is a large enough discrepancy in which case they may feel the need to take "corrective action."

For the piloting of the second release of the system, which is complete with an API to pass realtime data from an eLearning product, we made a plan which is based on these findings; to explore where they lead and to decide on the revisions to be made for the final release of the system.

3.2 PILOT STUDIES

The pilot took place at Maya Private School in Ankara where the language of instruction is English. 10 mentor teachers and 150 students of grades 7 to 11 participated in the 2 week study. The results are to be evaluated using TAM3 framework (see D5.4 Section 3.2) which demands a long enough piloting period for the participants to build up attitudes towards the technology product. During the 2 week time frame, students used the 12 Day Course Plan of a speed reading web application called HızlıGo, www.HızlıGo.com, developed by SEBIT. The application consists of 21 digital activities that target specific competencies on speed reading which are mapped to a structure of 35 competencies. As the activities are used the performance measures are extracted, translated and loaded to LEA's Box via its API in realtime where analytics tools are applied to form an Open Learner Model (OLM) for each user as well as their mentor teachers for the study. These OLMs and negotiation facility were made accessible through the LEA's Box portal. After the 12 Day Course of HızlıGo is completed with LEA's Box used for decision support, a concluding survey is administered online.

The below sections describe this pilot and the software, as well as the results which were compiled, analyzed and submitted as a paper for the ACM UMAP 2016 - 24th User Modeling, Adaptation and Personalization Conference.

3.2.1 HIZLIGO SPEEDREADING SOFTWARE

HızlıGo is a retail product developed by SEBIT. It compiles a set of activities to develop speed reading abilities and break certain bad habits that impede reading. It has a modern interface, a rich set of reports and an exercise section alongside the course plan.

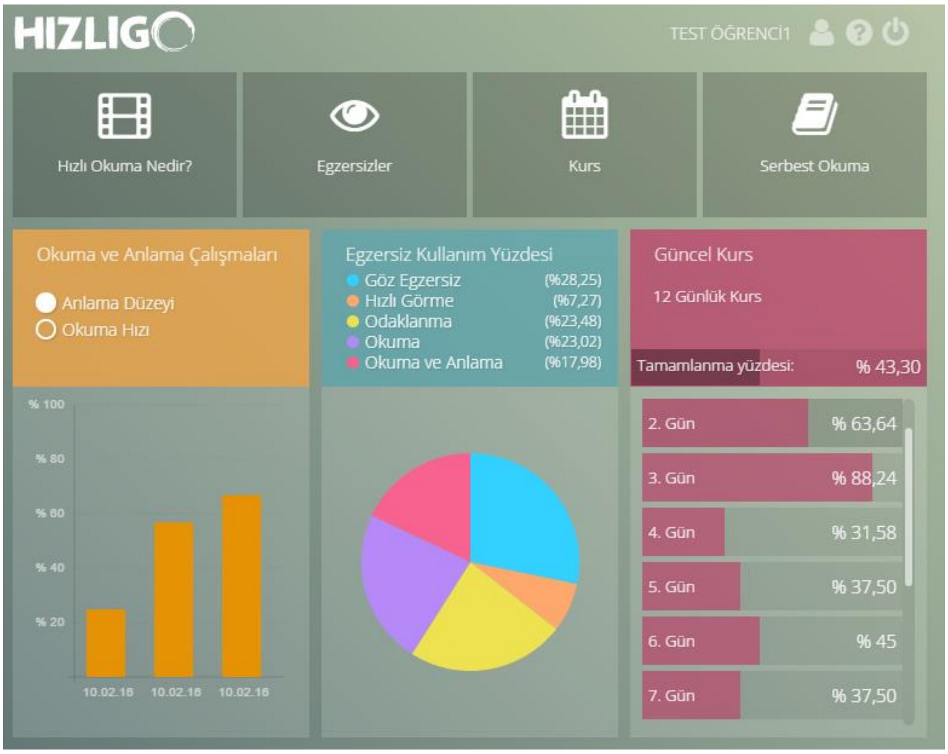


Figure 6. HızlıGo

The shortest course plan of HızlıGo was 21 Day Plan which was too long for the piloting resources available. Therefore a 12 Day course plan was created. HızlıGo was ideal for the pedagogical scenarios of this pilot because, those students who fall back or advance fast in the course plan were able to use the exercises section for remedial or additional work and they could use LEA's Box tools to decide which activities to do as an exercise.

To link HızlıGo to LEA's Box, the API defined at the second release of the system was used (see D2.2 Revised System Design Document). An adaptor software unit was developed to read in real-time the log data from HızlıGo, translate it to the format demanded by the API and call the API to load. This format necessitates adding user and activity information in the call. Therefore unique identifiers for each user is created and defined in the configuration tool of LEA's Box so that data LEA's Box could store the data without knowing the real identity of the user. Similarly activity information was codified, linked with the competency structure and defined into LEA's Box again using the configuration tool. Below is an excerpt from this competency-activity linking table.

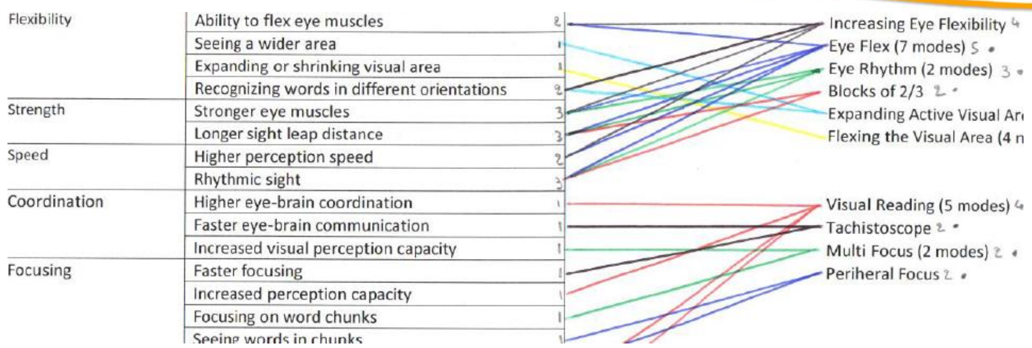


Figure 7. An excerpt from the competency-activity linking table

3.2.2 LEA'S BOX PORTAL AND OLM

As the data pours in to the LEA's Box portal in realtime from HızlıGo data adaptor, competency levels and states of each user is calculated and made available in many forms. New among these forms are the "Across Time" and "Heat Map" views. "Across Time" view displays the student progress in a competency (or in average) across time, hence revealing attendance information as well. Heat Map is a matrix that can display which activity contributes to which competency by how much, hence revealing the optimal activities that can be exercised to gain more.

HızlıGo has performance charts for each activity and in teacher accounts it is possible to compare students in pairs. Yet in LEA's Box OLM, all reports are competency-based and it is possible to observe all students and the class averages. HızlıGo demands a certain study period every day of the course plan and it reports what percent of each day's activities are completed. Yet in LEA's Box OLM, Across Time view displays exact times of attendance and how beneficial each attendance was.

Both teachers and student manuals are prepared and distributed for LEA's Box OLM as well as easy workflows for primary use cases.

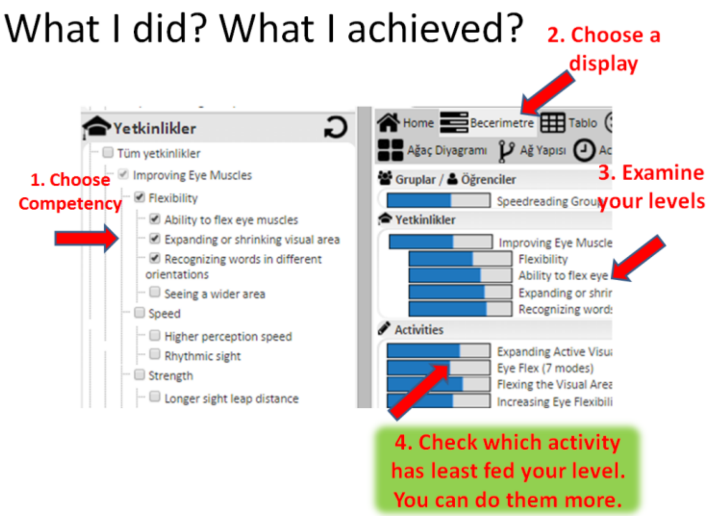


Figure 8. HızlıGo analyses

3.2.2 PILOTING

The study kicked-off on 10th February with an address to all participants in the conference hall of the school. In this seminar the concept of "competencies" in education was elaborated with metaphors such as "muscles" and difference between knowledge and competency was highlighted. LEA's Box was positioned as a decision support system, somewhat like a digital coach in training to develop competencies. Consent forms were collected right away by the teachers.

Figure 9. Piloting participants

On the 4th day of the HızlıGo course plan, there was enough data so the mentor teachers started to use OLM. Besides a manual and workflow presentation, videos from the LEA's BOX YouTube channel were used for teacher training. Once they were competent enough with OLM and could run basic usage scenarios, they introduced the toolset to their students. Basic use for students was to "login after each HızlıGo study day, check if they fall behind any competency, determine which exercise to do for closing the gap, and if they don't agree with their competency levels observe evidences and negotiate.

Students followed HızlıGo 12 Day course plan on their own, with occasional intervention by their mentor teachers who continually assessed their progress. Besides, 4 lab sessions were committed where reflection on evidence and use of negotiation facility were done 1-1 with researchers. Lab sessions tried to identify 4 inquiries:

1. What do the students understand from their model?
2. Can they identify their strengths / areas of improvement?
3. Do they agree with their model? Would they use persuasion to see evidence and update it?
4. Can they plan what to do next?

Figure 10. Piloting: Ali Türker presents the tools and ideas

3.3 Evaluation and Validation

We consider our analysis of this pilot in two parts: (1) the use of learner model persuasion as a method to address issues of model accuracy; (2) the validation and acceptance of the technology using the TAM3 framework.

3.3.1 Evaluation of learner model persuasion – reproduced from Johnson et. al (submitted)

Analysis of the results of this pilot that related to learner model persuasion are written up and submitted as a conference publication to User Modelling, Adaptation and Personalization (UMAP) 2016:

- Johnson, M.D., Ginon, B., Turker, A., Kickmeier-Rust, M., Kerly, A., Bull, S., Masci, D., Khurshid, A., Baber, C. (submitted). Supporting Stronger and Weaker Learners with Learner Model Persuasion when Evidence Comes From Third Party Eductional Tools. Submitted to User Modelling, Adapatation and Personalisation 2016.

Participants, Materials Methods

Whilst the majority of the method is covered earlier in the section, we provide a summary, to state that this analysis originates from: 59 learners aged 11-15 completed a 12 day speed reading course using Hizligo in Turkey, on a recommended basis of 30 minutes interaction per day. The LEA's BOX OLM was made available alongside their course. All use was optional and both tools could be localised to Turkish. All interaction in Hizligo resulted in an instant update of the OLM. On Day 1 of the course, students were introduced to both tools via a presentation and live demonstration. Several lab sessions were held during the course, where an expert user of the technology was present and students were able to ask questions and seek clarification. All usage of the OLM was logged and a 5-point Likert scale questionnaire about student use was administered on Day 12.

Results

We consider the use of learner model persuasion in the (a) wider context of use (b) competencies selected (c) amendments proposed and actions taken. In all cases, participants are categorised as 'strong' (29 learners) or 'weak' (30 learners) based on the final state of their learner model (above/below 49.8% - the median model state).

(a) Wider context of use

There was a good level of interaction (Table 1) with the external tool, and 46 episodes of persuasion are recorded (17 from weak learners; 29 from strong) from a total of 100 OLM accesses, by over half the sample. Weak learners were equally likely to use the OLM (15/30) or complete persuasions (8/15) as to not do so. Stronger learners were more likely to use OLM (20/29) or complete a persuasion (14/20), and overall had a greater level of engagement, potentially being more motivated learners. Of those that used persuasion, there is little distinction between stronger/weaker learners (average use 2.1).

		No. of Students	Range per student	Mean	Median	Total
Activities performed in external data source (Hizligo)	Overall	59	11 to 275	87.1	61	5140
	Weak	30	13 to 187	49.4	39.5	1482
	Strong	29	11 to 275	126.7	119	3675
Final OLM state (level, out of 100)	Overall	59	34.2 to 99.5	52.8	49.8	-
	Weak	30	34.2 to 49.8	43.6	43.9	-
	Strong	29	50.4 to 99.5	62.3	58.0	-
OLM sessions/logins	Overall	35	1 to 9	2.9	2	100
	Weak	15	1 to 9	2.6	2	40
	Strong	20	1 to 9	3.0	2	60
OLM persuasions initiated	Overall	22	1 to 9	2.1	1.50	46
	Weak	8	1 to 4	2.1	2	17
	Strong	14	1 to 9	2.1	1	29

Table 1: overview of activities performed, OLM state, logins and persuasions.

24 of 35 users of the OLM responded to the questionnaire (14 strong learners, 10 weak – Figure 11). Most users indicated they *understood* their learner model (16 agree, 6 neural, 3 disagree), and so stated that they could interpret the information. When identifying *strengths*, stronger learners were more able to do this (agree 8/14) than weaker learners (agree 4/10) and for identifying *areas of improvement* weaker learners were more able to do this (agree 7/10) than stronger (agree 3/14), which reflects students identifying the majority trend in the dataset, rather than outliers. Approximately half (agree 13/24) indicated using the OLM to think on a metacognitive aspect.

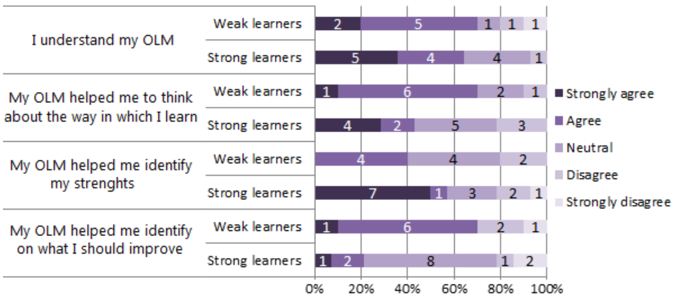


Figure 11: questionnaire items: understanding, cognition and metacognition

(b) Competencies selected for persuasion

The competency framework contains three grain sizes. From the 46 episodes of persuasion, the competencies selected at the coarse and medium grain sizes occur more frequently than is proportional to their occurrence in the 50–item framework (Figure 12). This is true to a greater extent for weaker learners than for stronger.

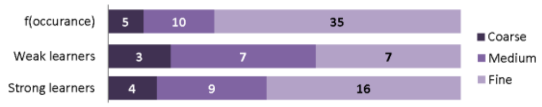


Figure 12: granularity of competencies that were persuaded

When considering the learner model state for each competency selected for persuasion (Figure 13), competencies that are identified as weak (49.8% or less – median value for all models) are slightly more likely to be selected than items that appear stronger. When each selected item is compared to the distribution of learner model of other competencies for the same student (threshold is the median value for the model), both stronger and weaker students are more likely to select competencies that appear as weak for their model.

Figure 13: state of competencies selected for persuasion, versus state of the model

(c) Amendments proposed and actions taken

A total 36 of the 46 episodes of persuasion included a self-assessment and a proposed change to the model (Table 2). All except two of the persuasions were to increase the level of the model (mean 16.1; median 11.5), with weaker learners proposing changes of greater magnitude (mean 21.7) than stronger learners (mean 12.6). Both requests to decrease the level of the model came from weaker learners.

Change to Competency		No. of Persuasions	Magnitude of Change		
			Range	Mean	Median
Overall	Increase	34	2 to 69	16.1	11.5
	Decrease	2	-17 to -10	-13.5	-13.5
Weak learners	Increase	13	2 to 69	21.7	16
	Decrease	2	-17 to -10	-13.5	-13.5
Strong learners	Increase	21	2 to 34	12.6	11
	Decrease	0	-	-	-

Table 2: direction and magnitude of proposed learner model changes (total: 36)

32 of the 46 persuasion episodes resolved to an agreed outcome (Table 3), with this being of equal likelihood for all learners (strong 20/29; weak 12/17; approximately 70% in both cases). Steps in the persuasion process (see Bull et al (in press) for full description) also allowed learners to request evidence from the model; the system justifying itself. This was done in half of the persuasions (23/46 episodes; 14/22 students), with equal likelihood for both strong and weak learners. With reference to the student providing justification to the system, this was completed with a greater frequency (18/22 episodes), with a total of 80 justifications being tendered. Weaker learners showed a greater engagement with this (8/17 learners; average of 4.6 justifications; median 3.5; minimum 2) as compared to stronger learners (10/29 learners; average of 4.2 justifications; median 2).

		No. of students	Range	Mean	Median	Total
Resolved (outcome)	Overall	20	1 to 8	1.5	1	32
	Weak	8	1 to 8	1.5	1	12
	Strong	12	1 to 3	1.7	1	20
Unresolved (outcome)	Overall	11	0 to 3	1.3	1	14
	Weak	3	1 to 3	1.7	1	5
	Strong	8	1 to 2	1.1	1	9
Provide Justifications (can complete multiple per persuasion)	Overall	18	1 to 29	3.6	2.5	80
	Weak	8	2 to 11	4.6	3.5	34
	Strong	10	1 to 29	4.2	2	46
Request Evidence (may only be done once per persuasion)	Overall	14	0 to 3	1.6	1	23
	Weak	7	1 to 3	1.6	1	11
	Strong	7	1 to 9	1.7	1	12

Table 3: persuasion actions taken (total: 14 strong learners and 8 weak learners)

Classification of the 80 justifications given by learners places them into three broad categories (Figure 14): (1) evidence not known by the system (24 instances) e.g. access of other resources, environmental factors; (2) potential erroneous evidence (32 instances) i.e. inconsistent, or data loss suspected; (3) discrepancies between the system and the student's point of view (24 instances) e.g. "I am more confident". Stronger learners were more likely to state omission evidence, whereas weaker learners more often stated discrepancies between the system and their understanding. Both were equally likely to state that they believed there were errors in the evidence.

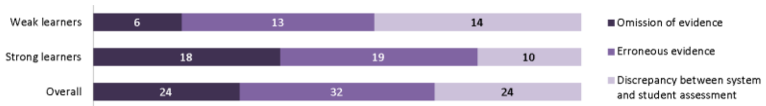


Figure 14: justifications tendered during persuasion: broad categories

Additionally, of the 15 questionnaire responses of the 22 users of persuasion (Figure 15), most students agreed that they used the persuasion feature to better understand the evidence behind their model and how it was calculated, in addition to being able to use persuasion to explain their point of view and make their model more accurate rank. In all cases, responses were similar between stronger and weaker learners.

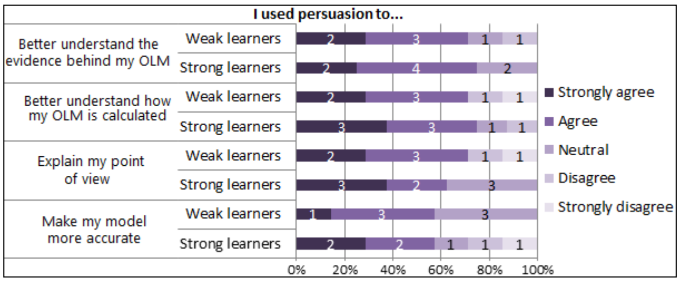


Figure 15: questionnaire items: reasons for using persuasion

Discussion

The students in our study, whether strong or weak learners, engaged with the concept of persuading their learner model, indicated that they understood it, and were equally likely to make use of it, regardless of ability. All use of the tool was optional, although as might be expected, stronger learners had a slightly higher level of engagement overall. Interestingly students used the visualisations in the learner model to identify a category of information that was present in the majority of the visualisation; that is to say stronger learners used it to identify mainly strengths, even though areas of improvement may have been the obvious outliers in the visualisation, and vice versa. This is an important consideration as the visualisations are the initiation point for learner model persuasion, and effecting changes in areas of potential inaccuracy.

When initiating persuasion students generally selected a competency showing as weak, relative to the rest of their learner model, and then almost always try to persuade its value up. In the case of weaker learners this is of a greater magnitude, and generally on a more coarse level of granularity, i.e. for a higher level competency in the framework. This suggests that particularly weaker learners may be trying to effect bigger changes at once on their model using the persuasion method, and also that the learners have an awareness of where areas of improvement exist, even if they don't intend to use the visualisations for this purpose.

Weaker learners were also able to use the persuasion feature equally as well as stronger learners to reach a resolution. Compared to stronger learners, they were more likely to add justifications and focus more on discrepancies in their viewpoint, rather than being as evidence based as with stronger learners. Arguably, reasons for using the persuasion function may not relate to addressing accuracy alone, but also providing a voice for the learner and as a means to understand more about what the information in their learner model means, and build a relationship with the underlying data, in the full context of interactions that have taken place with technology.

A key issue arises from this, and that is to consider how learner model persuasion might also judgement on competencies that are highlighted as strong, in terms of inconsistencies and errata. There is also a further issue that students would have been aware that whilst the information shown is formative, it can also be seen by the teacher. This behaviour is in alignment with students being protective of information that makes them appear strong, and justifying themselves in areas where they are not; a wider issue in assessment culture.

Conclusion

We have presented a use case for learner model persuasion in the context where the evidence from which the model is built comes from a connection with a third party system that: *can be seen as imperfect; can omit the full context of use; and may not fully correlate to students' perceptions of their learning experience.* In our evaluation students engaged with the LEA's BOX OLM which afforded them the option to persuade their learner model in cases where they identified that it should be updated. Our initial results have also indicated that persuasion may also have benefit in giving students a voice through which to explain their perceptions, and a means by which they may better understand the model, showing an interest in the underlying evidence, the learner modelling process, and self-assessing as a means to identify discrepancy. We present several recommendations and areas of consideration:

1. *Consider how different levels of granularity can be supported in persuasion:* learners may be attempting to effect more general changes on the learner model, particularly in the case of weaker learners. This raises a question regarding whether coarser grained persuasions might benefit from a different workflow and

more globally effective actions, above working on a competency-by-competency basis.

2. *Persuasion may help learners gain better understanding of how the OLM built.* Learners have an interest in information relating to evidence and modelling process.

3. *Learners may wish to use persuasion as a channel to voice their opinion.* Adding justifications and information regarding additional viewpoints and contextual evidence, even if not quantitative, may be useful for helping students build a relationship with the educational data, and to understand/think about the competencies more.

4. *Should stronger and weaker items in persuasion have different workflows?* Learners can identify areas of improvement in their learner model and use persuasion to try to increase the values of these. If this is a natural criteria for initiating an episode of persuasion, are different workflows needed to support thinking about strengths or areas of improvement?

5. *Persuasion, or other interactive maintenance, should be considered for inclusion where the learner model information source is third party.* Students do engage with this, and state it useful, and can use it to address issues of accuracy, omitted evidence, erroneous evidence and discrepancies between their experience and the OLM. Fundamentally this should be explored further in terms of user trust, and with the use of multiple systems. Educational data is now distributed over many different resources. Open learner modelling may yet have an important part to play bringing together information from many different origins and allowing students to see the overview of this information and consider its validity through the different actions available in learner model persuasion.

3.3.2 Validation using the TAM3 framework

Of the 160 students addressed, 87 attended HızlıGo and 59 did enough activities to form an opinion, as well as 10 teachers. The TAM3 survey collected from 55 students and 10 teachers. The data collected is placed in a correlation matrix. The values of correlations between variables and those of the partial correlations are compared by Kaiser-Meyer-Olkin (KMO) measure to reveal a 87% sampling adequacy. In statistics, KMO measure in 80s is considered to be “meritorious” to carry out factor analysis.

Factor analysis of the 5 variables that load “perceived usefulness” for LEA’s Box turned out to reveal the below order of influence:

Factor	Teacher (%)	Student (%)
System Quality	88	81
Relevance to job	85	80
Visible Results	89	79
Personal Image	92	78
Perceived Usefulness	**88**	**80**

Factor analysis of the 4 variables that load "perceived ease of use" for LEA's Box turned out to reveal the below order of influence:

Factor	Teacher (%)	Student (%)
Self-efficacy	82	89
Perceived control	82	92
Tech Anxiety	89	81
Playfulness	79	93
Perceived Ease of Use	**83**	**88**

These results validate that both students and teachers have a strong behavioral intention to use LEA's Box OLM tools. However, those intentions are hampered by mostly "perceived usefulness" factors in case of teachers, and mostly "perceived ease of use" factors in case of students. These results may be attributed to a general "belief" in students that it is the duty of the teachers and educational coordinators to decide about the "usefulness" of an educational technology and the conditions that they choose to adopt it or not mostly depends on how easy it is to incorporate that technology in their daily lives.

3.4 Summary of the benefits of LEA's Box OLM and further development

Learning analytics can provide a snapshot of what students know, what they should know, and what can be done to meet their academic needs. With LEA's Box, educators can make informed decisions that positively affect student outcomes. Research has shown that using data in instructional decisions can lead to improved student performance[1]. No single assessment can tell educators all they need to know to make well-informed instructional decisions, so researchers stress the use of multiple data sources[2]. Generally, enormous amounts of data can be collected on digital learning platforms regarding students' attendance, behavior, and performance, as well as administrative data and perceptual data from surveys and focus groups. But when it comes to improving instruction and learning, it's not the quantity of the data that counts, but how the information is used. Our piloting study proves that LEA's Box tools provide information for the teachers to make continual and formative assessment. Students can also benefit from being able to have more opportunities to demonstrate what they've learned, ask questions, and seek new knowledge about their learning. As for further development, there are already improvements made on the system based on piloting feedback to improve usability. For instance the Heat Map tool was updated to highlight those activities that combine to affect a competency with a bold black border.

Our observations for further development are as follows:

1 – Technical perfection and fluency at the first contact with the software are the largest determining factors for the users to participate the pilot and they may as well be so for adopting the toolset for everyday use. Factors such as an easy URL, easy login, simple use cases, browser support, mobile support affects hugely. This situation is also affected from peer influence. When students start to talk

[1] K. P. Boudett (2005), *Data Wise: A Step-by-Step Guide to Using Assessment Results to Improve Teaching And Learning*, Harvard Education Press

[2] L. Hamilton,et al. (2009). *Using student achievement data to support instructional decision making*, Institute of Education Sciences, U.S. Department of Education.

about the application being "cool" or being "cumbersome," the idea spreads very easily and becomes a general belief. In our case, there being two different systems that demand login seriously affected the participation ratios. This also means that when/if LEA's Box tools are integrated with products their adoption could be even higher.

2 – To benefit from LEA's Box tools, users should have a keen notion of what competencies are, especially compared to knowledge and skills. The simple tree structure competency filter menu in OLM is the main facility for a user to observe the competency structure. In case of speed reading competencies this menu was enough for users to grasp the competency set and its structure. However for complex, structures this may not be sufficient.

3 – Alarming situations such as one competency being abnormally low compared to others, or an array of evidences all being negative etc. should be more readily observable in the model. In addition, basic use cases such as which activity was least benefited from (given the competency), which students are fading etc. can be made more obvious.

4 – It takes a certain amount of discrepancy between the indicated competency level and the user's perceived competency level to initiate the negotiation facility. That means, negotiation (if committed) is more of an operation of correction than an operation of adjustment. There is also a usability issue in that, the competency to be negotiated is selected from a combo box rather than from the competency structure.

5 – Competency state of a student with respect to class and/or with respect to course plan are too great motivators to check on analytics results.

6 – There is a difference between "gaming" and "playing." Games have concrete rules; they are highly behaviorist and strengthen the particular skill set that the game is designed for. On the other hand, play is rule-free, risk-free, explorative, and constructivist in learning by trial & error. Learning Analytics tools should use more "playful" interface and interactivity paradigms, since the idea is to empower the user to draw individual inferences and surmises, have diagrammatic reasoning.

7 – Publishers and content developers could be very strong stakeholders as potential users of learning analytics to validate their products and underlying structures.

4. Piloting and evaluation in Austria (TU Graz)

4.1 Description of the Graz use case: Grazer Schulschwestern

In the following, we will describe the use case “Schulschwestern”, named after the according school in Graz where this use case will take place. The opportunity of making use of this case appeared just recently, at the end of January 2016, and that´s the reason why we haven´t mentioned this use case earlier, for example in one of the previous Deliverables in WP5 (D5.1 – D5.3). TUGraz and teachers and the heads of the Schulschwestern are still “negotiating” some details of the procedure, the material (i.e. questionnaires) and the set of tools to be applied.

During these “negotiations” so far, it became (again) obvious that there are some discrepancies between the aim to carry out scientifically-sound studies on the one side and the school reality, including obstacles, on the other side. Just as an example, it might not be possible to have a control group, due to the fact that the teacher cannot split his class into two parts (neither locally nor treatment-wise). We are currently trying to make the best out of such constraints and are exploring alternatives (for example: both halves of the class could switch their “role” of being part of the experimental or the control group in a second round of the study).

Due to the ongoing discussions, we can only describe the currently planned procedure – some details in the actual implementation of the use case study might differ from the description below.

4.1.1 Participants

A) Teacher
The teacher who will participate in this study completed his Master in Ecology and Evolutionary Biology in 2013, as well as an additional pedagogical training program. Since then he has been working as a Biology and Human Ecology and Resource Management teacher. The teacher also participated in the weSPOT project (http://wespot.net/) as part of a test-bed and is therefore already familiar with the weSPOT platform and its tools (see below).

B) Students
The students are around 15 to 16 years old. A single class consisting of 34 students will participate in the study.

4.1.2 Material

A) Toolset
The set of tools which will be used encompasses:

1. the weSPOT environment,
2. a tailored MyClass environment, including OLM,
3. the FCA tool,
4. the Flower app,

Ad 1.) The weSPOT environment

In the course of the weSPOT project (Working Environment with Social and Personal Open Tools), a platform which aims to support inquiry-based learning has been created (http://inquiry.wespot.net/). Inquiry-based learning (IBL) as weSPOT´s pedagogical backbone regards inquiries as learning experiences, in which students develop understandings of scientific ideas by engaging in research activities. In addition to this platform, an underlying pedagogical model has been developed. This model presents an inquiry as a process that encompasses six distinct though interconnected phases: Question/Hypothesis, Operationalization, Data Collection, Data Analysis, Interpretation/Discussion and Communication. Each phase includes a range of activities and tasks that require specific inquiry skills and support further development of these skills.

The basic weSPOT toolkit includes four core components: a web-based Inquiry Space engine (WIE), a domain knowledge representation component (FCA), a learning analytics component (LARAe) and a mobile component, PIM. A detailed description of these four components is provided in Table 4.

Table 4: weSPOT´s toolkit

Tool	Description
weSpot Inquiry space engine (WIE)	WIE is a web-based platform (a hub) developed by re-using and extending the open-source social networking framework Elgg (http://elgg.org/). WIE allows teachers and students to set up an inquiry project, organize and structure it according to their needs by activating selected components (widgets) from a broad range of those available per inquiry phase. Examples of widgets are: A Question widget, a Mind Map widget, a File upload widget, a Page widget, a Discussion Forum widget, etc. Next to widgets that are activated for a single phase, several tools are available throughout the whole inquiry, such as FCA and LARAe.
Formal Concept Analysis tool (FCA)	FCA is a domain representation and domain visualisation tool integrated in WIE. FCA allows structuring the learning domain using objects (i.e., files uploaded into the platform), attributes and learning resources (URLs).
Learning Analytics Reflection and Awareness environment (LARAe)	LARAe is a learning analytics tool integrated in WIE. LARAe provides an overview of all learners' activities in a particular inquiry and shows generated content at individual and group level.
Personal Inquiry Manager (PIM)	PIM is a mobile app that enables mobile access to the personal inquiry space in WIE. With this app, users can manage inquiries on a mobile device and add data as text or images to their personal inquiry spaces in WIE.

A screenshot of the main entry point of the weSPOTplatform, the weSPOT Inquiry space engine, is shown in Figure 16.

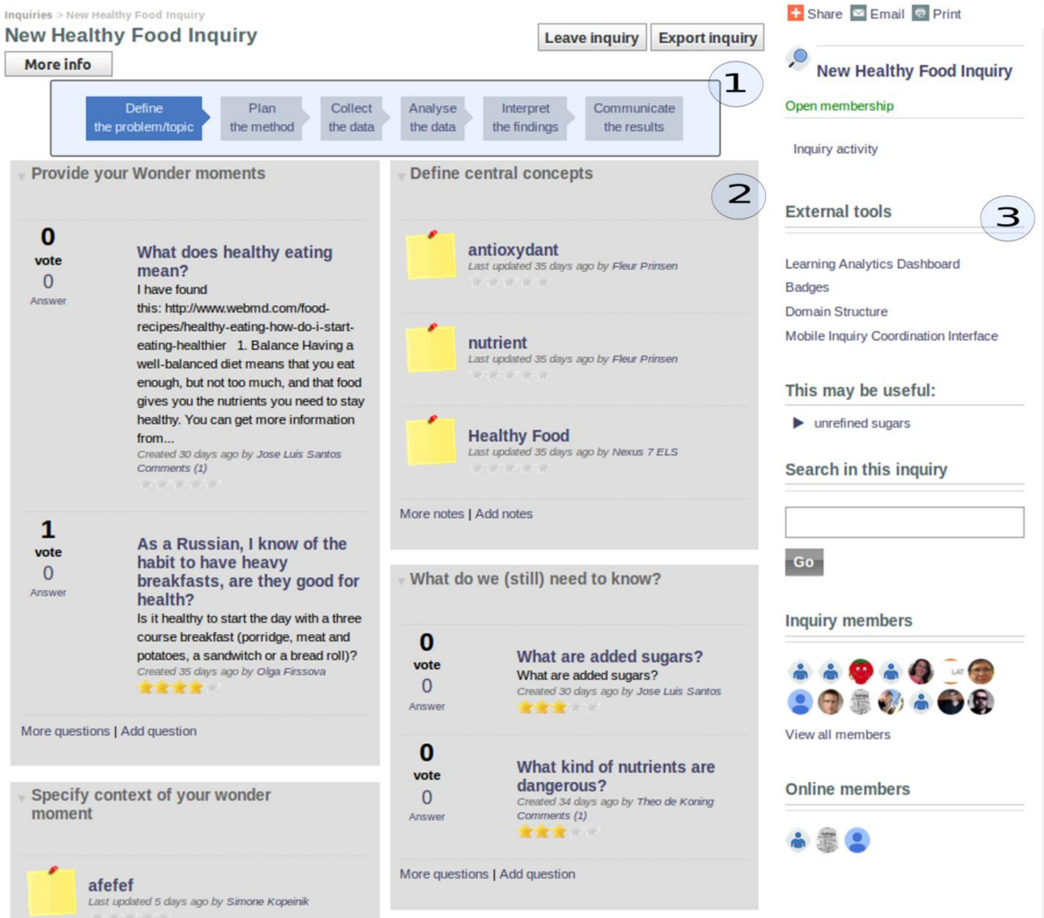


Figure 16. weSPOT Inquiry Workflow Engine with (1) the six inquiry phases represented as tabs, (2) the widgets space of the phase, (3) additional tools including learning analytics, FCA-domain structuring tool and mobile tool

Ad 2.) The MyClass environment

For a detailed description of the MyClass environment and its features see D2.2 (System Design Document II) and D2.5 (System Release II).

Ad 3.) The FCA tool

For an introduction to the FCA tool and its features see D2.2 (System Design Document II) and D2.5 (System Release II). Since the submission of this deliverable, a great deal of work has been put into new functionalities.

For example, the user can now switch between three different concept lattices (with different sets of objects and attributes) which enable to answer different pedagogical questions:

i) A concept lattice with learning activities as attributes and competences as objects give an overview of the learning domain and visualise the relations between competences and learning activities. By reading the concept lattice, one can easily detect which learning activities should be carried out in order to learn or train particular competence sets.

ii) A concept lattice with learners as attributes and competences as objects give an overview on the competence states of all students. An example is shown in Figure 17.

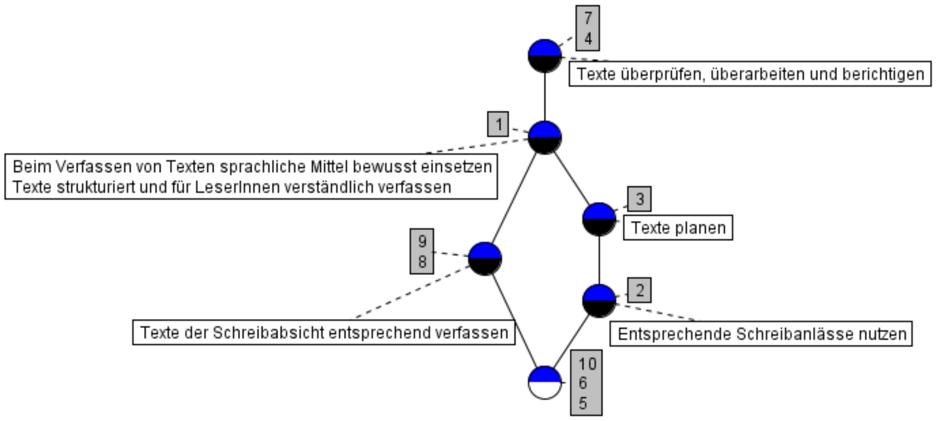


Figure 17. A concept lattice with students (digits, in grey boxes) and competences (in white boxes) represent the competence state of the whole class

Students are represented as digits (grey boxes). A particular students´ competence state can be identified by following all decreasing paths from the node where the student´s label is attached to and by "collecting" all competence labels in the white boxes. As an example, student 2 possesses the competence "Entsprechende Schreibanlässe nutzen" and the (current) competence state of the students 5, 6 and 10 is the empty set. Better performing students are located above lower performing students in the concept lattice, forming distinct groups and clusters.

iii)) A concept lattice with learners as attributes and learning activities as objects give an overview on the competence states of all students. An example is shown in Figure 18.

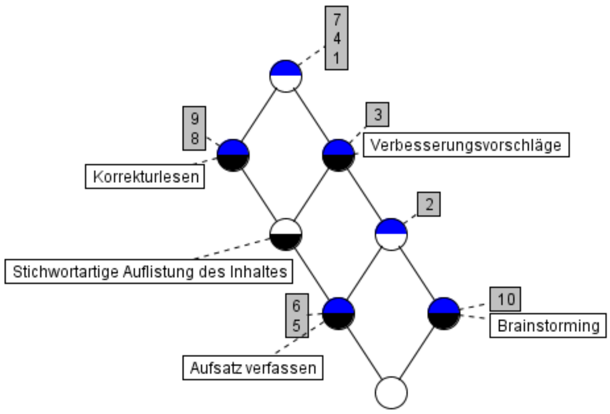


Figure 18. A concept lattice with students (digits, in grey boxes) and learning activities (in white boxes) represent the "performance state" of the whole class

The concept lattice shown in Figure 17 can be similar interpreted as the one shown in Figure : By following all decreasing paths from the node where the student´s label is attached to and by "collecting" all labels in the white boxes one can identify the learning activities which have been carried out by that student.

In general, this monitoring should facilitate teacher to support his or her students individually, by getting an overview on their strengths and weaknesses.

Ad 4.) The Flower Tool

The Flower Tool was described in the second chapter of this Deliverable.

B) Questionnaires and Inventories

1. Self-constructed knowledge test
2. Intrinsic Motivation Inventory
3. System Usability Scale
4. Semi-structured Interview

Ad 1.) Self-constructed knowledge test

To measure the students declarative knowledge states before and after using the intervention, i.e. the teacher uses LEA's BOX and the students use the weSPOT platform, a (domain-) knowledge test in line with the topic of the inquiry will be created by the teacher. In the ideal case, such a knowledge test should consist of around 40 multiple-choice items where only 1 out of 4 alternatives is correct. The whole set of items will be divided into two equally large tests, which should be as "parallel" (regarding difficulty, content, etc.) as possible. Those two parallel tests should have served as pre- and post-test. Half of the students should have received parallel test A as pre-test and parallel test B as post-test (and vice versa for the other half of students). This results in 20 items as pre-test and 20 as post-test.

Ad 2.) Intrinsic Motivation Inventory

To assess the intrinsic motivation of the students, we suggest applying the Intrinsic Motivation Inventory (IMI; Deci & Ryan, 2004; Ryan, 1982). From the original six subscales of the IMI, three subscales (interest/enjoyment, effort/importance, value/usefulness) were selected since they have been considered as most important for the purpose of this use case. Each subscale is represented by three statements, which have to be rated on a 7-point scale ranging from 1 (not at all true) to 7 (very true).

The IMI is provided in the appendix.

Ad 3.) System Usability Scale

The System Usability Scale (SUS) will be filled out by the teacher to assess his perceived usefulness and usability of the LEA's BOX tools. The SUS, developed by Brooke (1996), consists of 10 items. Items are formulated as statements (e.g.: "I thought the system was easy to use.") whereby the participant is asked to indicate the degree of agreement or disagreement on 5-point rating scale (ranging from 1 – "Strongly disagree" to 5 – "Strongly agree". Overall, the statements cover a variety of aspects of system usability, such as the need for support, training, and complexity. The score for each item is transformed such that it ranges between 0 and 4. The sum of all item-scores has to be multiplied by 2.5 and as a result, the SUS provides an overall usability and user satisfaction score having a range between 0 (negative) to 100 (positive). The SUS is provided in the appendix.

Ad 4.) Semi-structured Interview

The questions of the semi-structured interview are aiming to elicit the teacher's experiences with the LEA's BOX components, i.e. the MyClass environment and the FCA tool, i.e. the teacher´s perceived strengths, weaknesses and drawbacks of these components as well as his suggestions for improvements. The (quantitative) results of the SUS (see above) will serve as basis for prioritising the questions.

4.1.3 PROCEDURE

The overall procedure can be divided in three consecutive phases:

1. Phase: Pre-tests and Preparations

The students fill out the Intrinsic Motivation Inventory (IMI) as well as the first part of the declarative Knowledge Test before they first enter the weSPOT environment. These pre-tests serve as baseline and will be compared with the results of the consecutive tests in the remaining two phases.

The teacher sets up the inquiry in the weSPOT platform (e.g. the inquiry phases, widgets and the objects, attributes and learning resources via the weSPOT´s FCA tool). The topic of the inquiry will be related to the domain "applied biology". Together with members of TUGraz, the teacher will set up the LEA´s BOX platform, i.e. the students, the competences of the domain "applied biology" and the activities in the MyClass environment.

2. Phase: Intervention and Intermediate Test

The students interact with the weSPOT environment and work on their inquiries, partly in groups and partly individually. For this, they can use available tablet PCs. Their inquiry-related activities will be tracked by weSPOT´s LARAe tool (see section 4.1.2 above).

In the middle stage of the inquiry project (presumably after around 3 weeks) the students fill out the IMI for the second time.

The teacher uses the MyClass environment to manually track the student's activities, their competence-centred strengths and weaknesses during the lessons (as a kind of digital class book). In addition to that, he uses the FCA tool to gets an overview of all students of the class with regards to their competences and their competence-related learning activities. This monitoring should facilitate him to support his students individually, by getting an overview on their strengths and weaknesses.

3. Phase: Post-tests

Once **the students** finish their inquiry project (presumably after around 6 weeks), they fill out the IMI for the third time and the second part of the declarative Knowledge Test. Finally, the students evaluate the extent to which they possess the competences, predefined by the teacher (see above). This self-assessments will be compared with the teachers assessment and the comparisons will serve as basis for a negotiated Open Learner Model.

Kommentar [MB1]: Genau das ist ja mit den negotiated OLMs gemeint, oder?

The teacher fills out the SUS twice: First with regard of the MyClass environment and second with regard of the FCA tool. The results of these quantitative measures will structure and prioritize the topics of the concluding semi-structured Interview which will be carried out from a member of TUGraz.

4. Analysis

The teacher explicitly shares our interest for the question if and to which extend there are overlaps between the students self-assessment, the teachers assessment and the “systems” assessment. By “systems” assessment we mean the automated tracking via the LARAe component of the weSPOT platform. To analyse this question a rather simple correlational design can be applied.

The second main question is the change of the student’s intrinsic motivation and declarative knowledge over time. With regard to both variables we would expect an increase. These hypothesises can be tested by an by a Multivariate Analysis of Variance, MANOVA, (with e.g. gender as between subject factor and time of measurement as within subject factor and intrinsic motivation and declarative knowledge as dependent variables). Finally, we are interested in the teachers experiences with the LEA’s BOX components, i.e. the Next-tell based environment and the FCA tool. For this we will mainly focus on a qualitative analysis of the semi-structured interview.

4.2 Pilot: „GHS in der Taus“, Germany

As described in previous deliverables, the Taus school in Backnang, Germany is a so-called „Gemeinschaftsschulen“ [collective school] where all students from the 1st to 10th grade attend 1 common school form. Specifically the teachers of the Taus school are dedicated to this school form and developed a formative, competency-focused pedagogical concept at an early stage, which relies to a certain extent on Learning Analytics techniques and tools. The primary level comprises in total 628 students.

Figure 19. Photo from Taus school, in Backnang, Germany.

Although the pedagogical concept of the school is quite innovative, a problem was in the first years of our cooperation (which already started in the Next-Tell project) was the technical infrastructure. Robust Wifi access was lacking as well as the necessary technical dives for teachers. In the past year, this infrastructure was extended and now tablet computers and high-speed internet access is broadly available. The funding for modernization primarily came from the community of Backnang.

This boost in technical infrastructure also was a boost in the uptake of tools from Lea’s Box. As described in previous deliverables, this school is primarily using a myClass version (see the following screenshots).

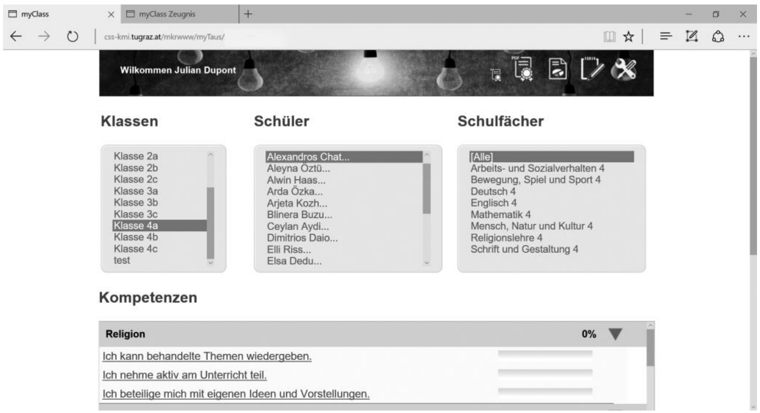


Figure 20. Screenshot of the main interface of myClass for this school.

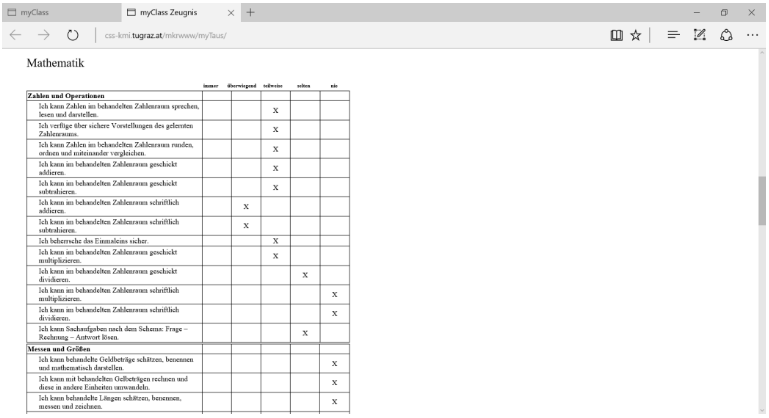


Figure 21. Screenshot of a typical, competence-based learning progress report.

The tool was used on a frequent basis by the teachers of the primary level. Primarily, teachers used the myClass tool through the portal for the frequent progress tracking of students' achievements. In monthly meetings, the teachers discussed the value and possible functional improvements. On January 28, 2016 we also a face-to-face workshop in Backnang. In the context of this cooperation and the design iterations, we developed a next level of design recommendations and feature lists for future developments (the details are given in Annex 3).

While in the past the tools were used in the primary level classes only, for the next school year (2016/17) it is planned to use the tools and the Lea's Box platform for the secondary level as well. In the summer term 2016 an introductory phase of this pilot is planned. Since the secondary level is conceptually different, modifications to the features and to the reporting of learning progress were necessary. Therefore, primary school teachers were included on a frequent basis to co-design the relevant features (the details are given in Annex 4).

To retrieve some more concrete figures we issued a short, 10 item survey (7 step scale transposed to percentage values for the follow chart) to teachers:

In comparison to the situation before using Lea's Box tools, Lea's Box:

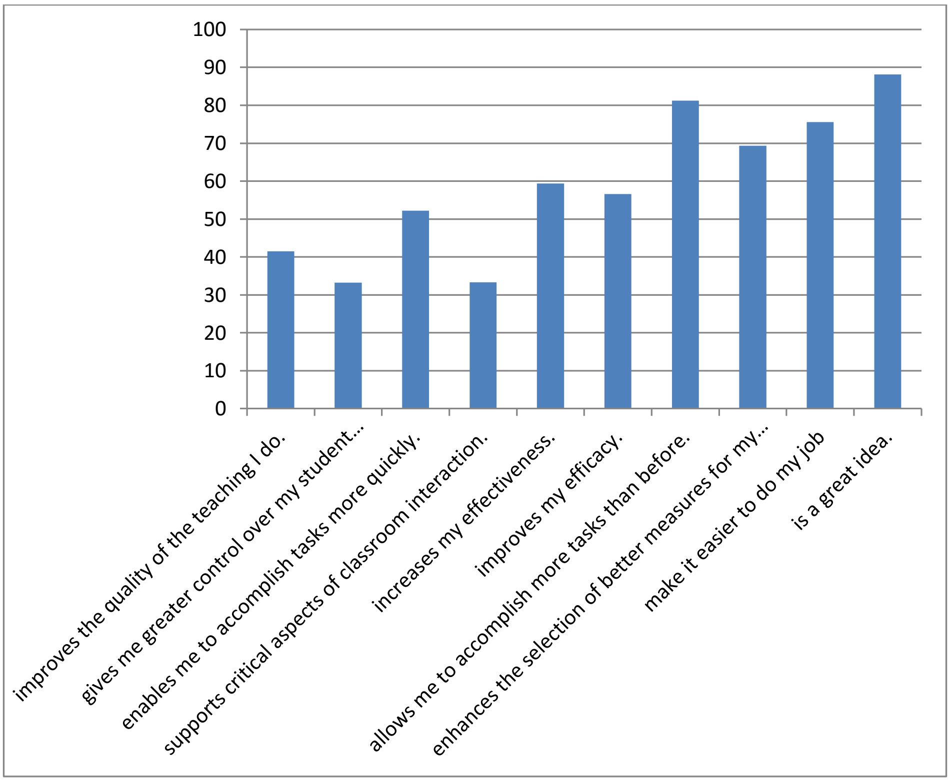


Figure 22: Results of a short TAM survey

4.3 Pilot: „KPH Graz", Austria

As already reported, the practice primary school of the „Kirchlich Pedagogische Hochschule Graz" [Catholic Teacher Training Academy Graz] is a private school under public law including after school care. The school has a strong focus and a school partnership where teachers, students, and parents jointly plan and build the school day and the setup of the school year. The school tries to actively live strong elements of 'progressive education' such as Montessori or Jenaplan pedagogy. Insofar, a broad tracking of activities and achievements is vital. Primarily, teacher use their version of myClass (as introduced in previous reports) for activity tracking the reporting of learning progress.

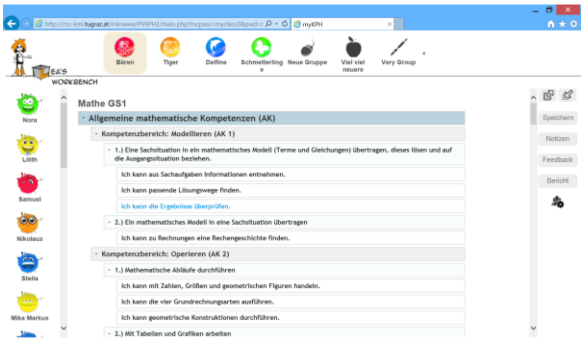


Figure 23. Screenshot from the myClass version for the KPH

On monthly basis we had meeting with a core team of teacher discussion the situation, the needs, and further developments. One major advance was to adjust the Lea's Box configuration module to the needs and wishes of the KPH teachers (see screenshot).

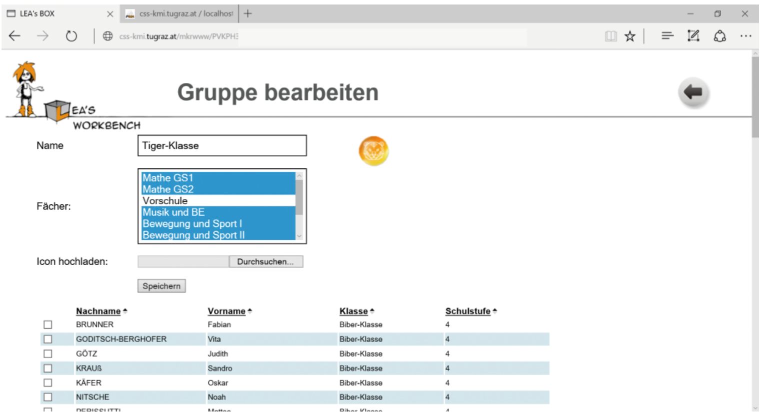


Figure 24. Configuration tool for the KPH version of myClass.

To retrieve some more concrete figures we issued a short, 10 item survey (7 step scale transposed to percentage values for the follow chart) to teachers:

In comparison to the situation before using Lea's Box tools, Lea's Box:

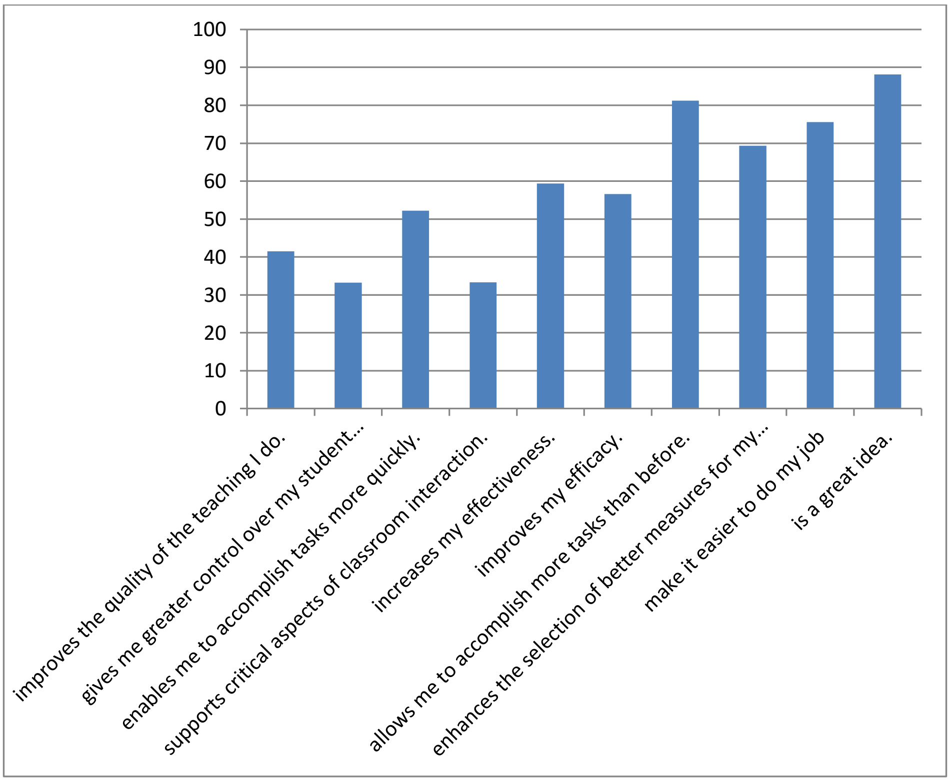


Figure 25: Results of a short TAM survey

4.4 Technical Evaluation: Adaptive Learning Research Group, Brno

In the context of the project we got in touch with the Adaptive Learning Research Group in Brno, which is a working group of the local University (http://www.fi.muni.cz/adaptivelearning). During the Lea's Box summer camp in Prague, we established a lively discussion and ongoing cooperation. A common ground is the fact that the research group in Brno, as well as TU Graz have developed learning apps, more specifically adaptive learning apps, for example in the field of mathematics.

The interesting thing is that although these apps cover the (almost) identical sets of competencies, they provide different types of adaptations and feedback to the learner as well as different user interfaces. The following shows the example of two multiplication apps.

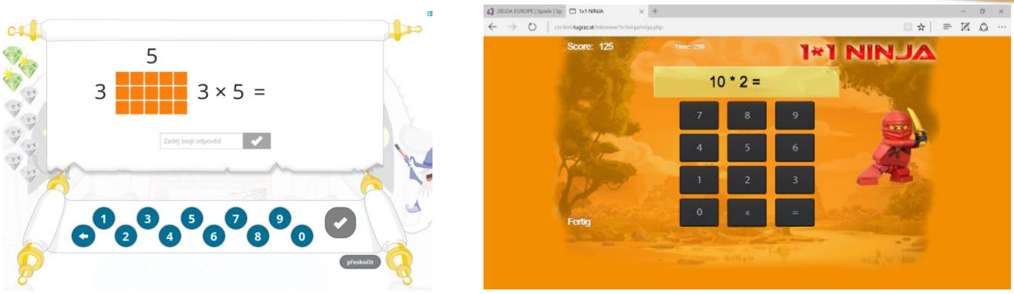


Figure 26. The left image shows the multiplication app from Brno, the left the 1x1 Ninja from TU Graz.

This cooperation pursues 2 strands, the one is the research-oriented strand, which aims at looking into the possibilities of CbKST and FCA to model different forms of analyses and subsequent adaptations. Results will be reported in the context of WP2. The second strand, that is important for WP5, is to investigate and explore the possibilities of linking external tools, such as the learning apps of Brno to the central Lea's Box platform. The concrete work was started at a face to face meeting in Graz on December 14, 2015.

The first concrete steps provide evidence that the integrated API of LEA's BOX is an easy and flexible method to link external tools into the system, without requiring extensive efforts of writing data adaptors. In the following project period, we are exploring standards such as xAPI/TinCan.

5. PILOTING AND EVALUATION IN THE UNITED KINGDOM (UOB)

Evaluation the United Kingdom has been completed as part of ongoing work that is contributing to the field of open learner modeling (OLM) and the interactive maintenance of OLMs. Engagement with university level students has been sought to further refine some of the concepts involved in learner model negotiation and learner model persuasion, in advance of evaluation with younger age groups. Many examples of the state-of-the-art in this field are in Higher Education, and so we have first aimed to validate our approach in this context.

5.1 LEARNER MODEL PERSUASION IN THE CONTEXT OF A UNIVERSITY LANGUAGE COURSE

We report here initial work students of Italian as a foreign language in terms of their initial impression, engagement and feedback from a focus group held. The outcomes of this have been written up as two conference publications, and we report here on themes and content discussed in the following papers:

- Bull, S., Ginon, B., Boscolo, C., and Johnson, M.D. (in press). Introduction of Learning Visualisations and Metacognitive Support in a Persuadable Open Learner Model. To appear in the proceedings of Learning Analytics and Knowledge 2016, Edinburgh, UK, April 2016.
- Ginon, B., Boscolo, C., Johnson, M.D., Bull, S., (submitted). Persuading an Open Learner Model in the Context of a University Corse: An Exploratory Study. Submitted to Intelligent Tutoring Systems, 2016.

5.1.1 PARTICIPANTS, MATERIALS, METHODS

Participants are 52 volunteer undergraduate students in taking a course in Italian language, at the University of Birmingham, UK. All received a demonstration of the LEA's BOX OLM and several drop in sessions took place with expert users of the OLM present to provide assistance to learners. Students complete quizzes in their content management system, Canvas, and the results of these are imported into LEA's BOX on a daily basis, as the evidence from which the learner model is built. These are combined in the model with teacher assessments at periodic intervals. At the start of the course 5-point Likert scale questionnaires were administered to identify students' anticipated use of the technology. At the end of month two further 5-point Likert questionnaires were administered with regard to use of the persuasion feature in the technology, and a focus group was held with 5 participants from the sample. All interactions are logged, and the study is due to complete in May 2016.

5.1.2 STUDENTS' INTENTIONS – REPRODUCED FROM BULL ET. AL (IN PRESS)

A total of 25 students responded to the questionnaire, which focused upon (1) intention to use visualisations, (2) how they expected to use the visualisations, (3) expectations about persuasion.

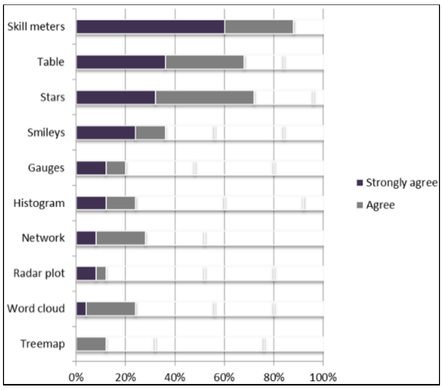


Figure 27. Expected use of visualisations.

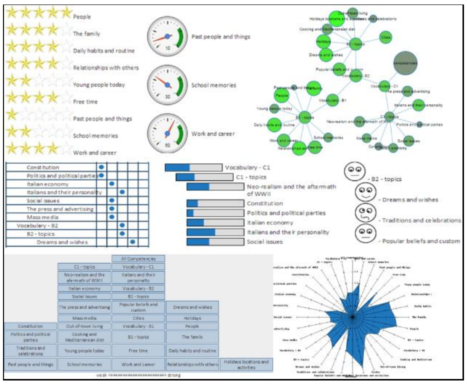
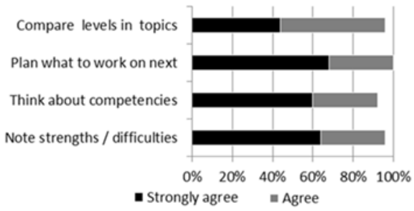


Figure 28. Visualisation types.

Of the visualisations (Figure 27) the skill meters are anticipated to be the most likely used, followed by the table and stars (Figure 28). The radar plot and treemap are expected to be least used. Most students indicated that they intend to use several visualisations: mean 3.84 visualisations; median 4; range 0-10. Figure 29 shows participants' stated expected purposes for accessing OLM visualisations. All purposes (comparing topics, planning, reflection, identifying relative strengths and gaps) are expected to be highly relevant. 23 of 25 students gave positive responses for all four purposes of viewing their learner model; 1 gave positive responses for 3 purposes (omitting the reflection option); and 1 indicated that they would use the OLM for only one given purposes (planning).



Purposes	Mean	Median	Range
Compare levels in topics	4.4	4	3-5
Plan	4.68	5	4-5
Think about competencies	4.52	5	2-5
Identify strengths/difficulties	4.6	5	3-5

Figure 29. Students' stated purposes for viewing the OLM

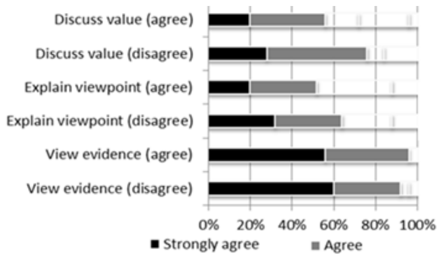


Figure 30. Students' expectations for using discussion feature.

Figure 30 shows participants' expectations regarding their use of the discussion component of the LEA's Box OLM. 23 of the 25 students claim that they would want to view the evidence for values when they disagree with them; and 24, when they agree with the values. 19 expect to discuss values when they disagree with them; and 14, when they agree. 16 stated that they wish to explain their viewpoint (justify their self-assessments) when they disagree with values; and 13, when they agree.

While some values are lower, the medians show that most participants expect to engage in discussion with the system, regardless of whether they agree or disagree with the values shown in the OLM.

We summarise our findings with 7 implications/recommendations for designers of OLM technology:

1. *As well as any visualisations that are particularly relevant to the specific context, include simpler visualisations such as skill meters or similar displays to help students identify a visualisation they can envisage using.*
2. *Offer multiple learning visualisations in an OLM to allow students to identify a range of options that they consider suitable.*
3. *Offer at least one structured visualisation in an OLM.*
4. *Explain how an OLM can support metacognition and self-regulated learning to ensure that learners are aware of this purpose.*
5. *Offer evidence for learner model values, as a means to facilitate self-monitoring, reflection, planning, etc.*
6. *Offer provision for learners to justify their own viewpoints on their understanding, skills, etc., as a means to further prompt metacognitive processes, even if an open learner model does not have a persuade or negotiate facility.*
7. *Allow students to discuss the contents of their learner model with the system if this is feasible in the context of use (i.e. if the learner modelling is sufficient to be able to support this).*

5.1.3 Initial response to persuading the learner model – reproduced from Ginon et. al (submitted)

Students were asked to complete a 5-point Likert scale questionnaire about their first impressions after 2 months' use of the OLM. The questionnaire, regarding the use, or non-use of the persuasion feature was responded to by 11 students (Figure 31); (3 used, 8 did not use). All students who did not use persuasion agreed with their competency levels (5 strongly agree, 3 agree), and all 8 indicated that they understood it. Only 1 student indicated they refrained because it was not summitively assessed. Conversely, all students who used persuasion indicated they disagreed with their model, wished to make it more accurate (2 strongly agree, 1 agree) and wished to explain their point of view (1 strongly agree, 2 agree). All 3 indicated that they wished to understand the evidence behind the model.

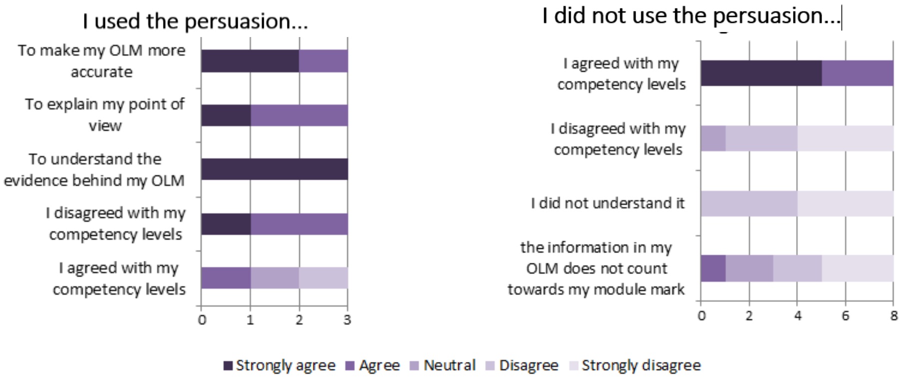


Figure 31. Students' reasons to use or not the persuasion feature

Semi-structured interviews took place with a sample of 5 of the students after 2 months of use, and focused upon their perceptions and attitudes towards learner model persuasion in the context of their learning. Students were asked whether they had used the persuasion feature, why they might use it and also why they might not make use of it. Students talked around their opinions and perceptions with regard to these points, in dialogue with the researcher. Students have their LEA's Box OLM open in front of them in a web browser. Interviews were recorded, then transcripted, and each lasted between 10-20 minutes.

Transcripts from the interviews of 5 participants showed student reasons for using (or not using) persuasion as fitting into several main categories, relating to information, time, precision and attitudes (see Table 5). To retain an accurate model, persuasion might be needed where students are completing short term goals (e.g. part completion of quizzes, leading to the model showing a lesser competency), because of limitations in the inferencing technology (e.g. usability problems, incorrect marking or multiple right answers, long quizzes) or because of more transient constraints (e.g. out of time and so the quiz is assessed too early). Students also indicated that persuasion of the model might not be a priority for them because they already perceived their model as accurate, because they hadn't covered enough of the course content, or because it wasn't at the point of the course during which it would be of the most use (namely this was stated as being during the summer exam period.).

Table 5: Themes occurring during interviews with 5 students about use of the persuasion. Numbers in [brackets] indicate the number of participants who talked on this issue.

	Persuasion	No Persuasion
Informational	**[5]** Course is large. Model needs more info. **[3]** Only completed part of the quiz, as this is most relevant => evidence is inaccurate. [2] Restore to a previous state. New quiz information is not representative. (Ctrl+Z) [2] Quiz content was too broad. Not everything is covered.	[2] More info is required first. [2] Difficult to associate numbers with persuasion.
Temporal	**[3]** Quizzes take a long time to complete. **[3]** Student out of time, quiz submitted early. [1] Wanted immediate feedback, quiz not complete, showing student as not competent.	[2] Takes time to complete persuasion.
Inference Precision and Level of Interaction	**[3]** Part of sentence not typed, answer still correct, but quiz marked it as incorrect. **[3]** No half marks for part-correct. e.g. case sensitive [2] Setup error causes incorrect marking. [2] Quiz platform is slow and not touchscreen => little interaction => OLM doesn't show current competency. [1] Right answers placed in wrong boxes.	**[4]** Don't feel have done enough quizzes yet use persuasion most effectively. **[3]** Already accurate.
Attitudes and Personal Learning Strategies	[1] Learning strategy leads to weaker inferences, e.g. use of trial and error	[2] Not exam period [2] Not confident with technology. [1] Not summatively assessed

Learner models are designed to represent learners' current skills, knowledge, competencies, etc. Usually they are assumed to be as accurate as is necessary for the purpose of personalising teaching. Within our small sample, our aim was to explore the use or otherwise, and reasons to use or not use a learner model persuasion feature:

1. *The model is already accurate or interaction is not time critical.* Within our sample, many of the participants said that they agreed with their model, and so there was no reason to try to update it. Some stated that it was perhaps the wrong time in their learning to use persuasion and may wait until an upcoming summative assessment before more intense engagement with some aspects of their course.
2. *Persuasion may help students understand better how the model is built.* Of those who claimed to have started model persuasion, each had a keen interest in seeing the evidence behind their learner model. This may suggest that a core foundation to persuading the learner model is to understand precisely where the information has come from, when it was generated, and for the learner to think about the differences and similarities between this and how accurate they perceive their learner model to be.
3. *Awareness of the limitations of the connections between technologies is a case for persuasion.* Participants showed awareness of some limitations of the quiz engine, such as

stringent scoring if they forget to add the end to a sentence but the answer is still correct, putting an answer in the wrong box, errors in quizzes, or using quizzes with their own learning strategies (e.g. choosing to work on only small parts of the course content at a time). In these instances, the values sent to the learner model will underestimate abilities, and this is a case for persuasion.

Some of our findings may generalise to other contexts – for example, university students appear to understand how learner model persuasion applies to their learning, when it might be useful, and are willing to challenge evidence when they disagree, explaining their point of view to the system. Such persuasion will allow them opportunities to try to influence the model content, giving them more control over their learning in, for example, an intelligent tutoring system where teaching is personalised according to the learner model. This control will will be further increased as we develop the persuasion to a full negotiated learner model.

6. SUMMARY

Deliverable 5.3 stated the following: "*We believe that for further successful development of the project it is necessary to modify the existing tools so as to be a less demanding for teachers, while displaying a clearly evident reason for their use. This will make it possible to increase the motivation of teachers to use them and help the future successful dissemination of the project outcomes*." This became the basis of our work when planning and conducting pilot activities of stage 2.

As we wanted to display a clearly evident reason for the use of the LEA's Box tools, as stated above, we needed to develop the tools in such a way so as to respect what teachers and students need most. That's why we took some time to re-analyze the feedback from the first stage of piloting before proceeding to the second stage and integrated the findings into our activities – for example, in the Czech Republic, we found teachers need suitable tools for self-assessment, so we included this into the so-called Flower-Tool.

We also took into account the need to make the tools as user-friendly as possible, and that's why all the pilot studies included questionnaires or other ways to get feedback regarding the user-friendliness of the tools, the most suitable visualizations of the results or other factors that may have influence over the probability the tools will be used in the future.

In general, the feedback regarding the pilot studies and use-cases was very positive and we dare say that in several aspects we saw a significant improvement over the feedback received after the first stage of piloting, especially when it comes to the perceived usefulness of the tools and their user-friendliness.

In conclusion, we are on the right track, but of course, developing the right tools is a never-ending process, and that's why we are going to keep improving and updating the tools based on the feedback we are continuously receiving, such as the interviews with teachers that are about to take place in the Czech Republic, and based on the results of new use-cases, such as the one that is about to take place in Austria.

References

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of educational psychology*, *84*(3), 261.

Brooke, J. (1996). SUS: a „quick and dirty“ usability scale. In P.W. Jordan, B. Thomas, B.A. Weerdmeester & A. L. McClelland (Eds.).Usability evaluation in industry (pp. 189-194). London: Taylor & Francis.

Bull, S., Ginon, B., Boscolo, C., and Johnson, M.D. (in press). Introduction of Learning Visualisations and Metacognitive Support in a Persuadable Open Learner Model. To appear in the proceedings of *Learning Analytics and Knowledge* 2016, Edinburgh, UK, April 2016.

Coury, B. G., Boulette, M. D., & Smith, R. A. (1989). Effect of uncertainty and diagnosticity on classification of multidimensional data with integral and separable displays of system status. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *31*(5), 551-569.

Deci, E. L., & Ryan, R. M. (2004). Intrinsic Motivation Inventory [Online]. http://www.psych.rochester.edu/SDT/measures/IMI_scales.php

Ginon, B., Boscolo, C., Johnson, M.D., Bull, S., (submitted). Persuading an Open Learner Model in the Context of a University Corse: *An Exploratory Study. Submitted to Intelligent Tutoring Systems*, 2016.

Johnson, M.D., Ginon, B., Turker, A., Kickmeier-Rust, M., Kerly, A., Bull, S., Masci, D., Khurshid, A., Baber, C. (submitted). Supporting Stronger and Weaker Learners with Learner Model Persuasion when Evidence Comes From Third Party Eductional Tools. Submitted to *User Modelling, Adapatation and Personalisation* 2016.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. Journal of Personality and Social Psychology, 43, 450-461.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, *4*(4), 295-312.

APPENDICES

ANNEX 1: INTRINSIC MOTIVATION INVENTORY

Dear Participant,

The following statements concern with your experience with the task you just engaged with. For each statement, please indicate how true it is for you, using the scale from 1 to 7. A 1 indicates that the statement is not at all true for you - with a 7 you indicate that the statement is very true for you.

	Not at all true					Very true	
1. I thought this was a boring task.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
2. I think that working on this task could be useful.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
3. I tried very hard to do well at this activity.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
4. This task was fun to do.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
5. I believe working on that task could be beneficial to me.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
6. It was important to me to do well at this task.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
7. I would describe this task as very interesting.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
8. I believe working on this task could be of some value for me.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
9. I put a lot of effort into this.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

APPENDIX 2: SYSTEM USABILITY SCALE

	Strongly Disagree				Strongly Agree
I think that I would like to use this system frequently.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
I found the system unnecessarily complex.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
I thought the system was easy to use.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
I think that I would need the support of a technical person to be able to use this system.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
I found the various functions in this system were well integrated.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
I thought there was too much inconsistency in this system.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
I would imagine that most people would learn to use this system very quickly.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
I found the system very cumbersome to use.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
I felt very confident using the system.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
I needed to learn a lot of things before I could get going with this system.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

ANNEX 3: INTRINSIC MOTIVATION INVENTORY

ANNEX 4: INTRINSIC MOTIVATION INVENTORY