Deliverable D5.6:

Piloting and evaluation report 3

Document Version:	December 23, 2016
Document Status:	Version 9
Document Type:	Report
Diss. Level:	Public
Lead Partner:	SCIO
Delivery Date:	M34 (December 2016)
Authors:	Andrej Novik (SCIO) Lenka Firtova (SCIO) Jan Synek (SCIO) Ludmila Miklušová (SCIO) Eliška Sovová (SCIO) Kristýna Fischerová (SCIO) Markéta Majerová (SCIO) Mustafa Ali Türker (SEBIT) Sezin Alkabir (SEBIT) Türker Özdamar (SEBIT) Gözde Kansu (SEBIT) Michael Kickmeier-Rust (TU Graz) Michael Bedek (TU Graz)

TABLE OF CONTENTS

1. INTRODUCTION

Deliverable 5.6 presents the third and the last stage of piloting of the Lea's Box tools and it covers the activities done during the last 9 months of the project, specifically from 1st of March 2016 until the end of the year. Being the final stage of the project, relevant documents in this busy period were:

- The DoW, the Annex to the D1.2 Periodic Report of Year 2: This Annex presented an evaluation framework which structured and prioritized the piloting and evaluation studies for Year3, steering them more towards exploitation opportunities. It also included a draft exploitation plan.
- D6.5 Dissemination and Exploitation Plan: Delivered at the end of August, this deliverable categorized the exploitation activities ensued by piloting and evaluation studies into those that involved the whole system and those that involved specific business cases that demanded specific tools in the system. It listed the anticipated actions for the following 3 months.
- D3.4 Third Release of LA/EDM Services and Algorithms: In September this deliverable has documented the most recent tools and services included in Lea's Box.
- D2.6 Final System Release: Delivered in conjunction with D3.4, this is an inventory of all the tools and services that are included in Lea's Box, and the API with which they can be accessed and employed.

The main objective of piloting and evaluation activities was to use the Lea's Box system and its components in real schools, to evaluate key aspects of the system and subsequently to modify the tools based on the feedback received.

Earlier deliverables from work package five were very practically oriented, for example D5,4 presented different educational scenarios and classroom issues, while D5.5, described specific tools offer by the Lea's Box system. Based on the feedback from received in year 1 and 2, the final release of the system was shaped and the integration of its individual components was considerably improved. This deliverable thus presents the main results of piloting and evaluation studies related to the whole Lea's Box system.

Activities described in this deliverable consist of several interrelated steps: Co-design activities, Pilots and Evaluation.

The first step entails the development of suitable tools in cooperation with end-users. In this step, we have always tried to engage teachers as much as possible, so that the resulting tools would meet their real needs and would be perceived as beneficial.

The second step, piloting, can be characterized as the deployment of tools developed in step one in real classrooms with the aim of answering one or more pedagogical questions outlined in the Annex.

The final step, evaluation, is done with the aim to assess the impact and suitability of the tools designed and piloted eariler. The main aspect that were evaluated were defined in the General Evaluation Matrix, as mentioned above.

In response to the Y2 review we sketched a global piloting and evaluation framework. This framework identifies the relevant system components and specifies the dimensions for evaluation and piloting. For this matrix we proposed priorities for the evaluation activities. Insofar, this framework provides one benchmark for the activities of WP5 across the entire project duration. In the following sections we report on the piloting and evaluation activities and link them to this matrix.

Lea's Box piloting and evaluation framework.

	Portal and Appr.	myClass	myActivities	FCA Tool	Mindmap	Learning Spaces	Hasse Diagrams	OLM	Config Tool	Flower App	Interconnection between tools and overall functionality	Rank
Functionality and adequacy	+++	+++	-	+++	++	+++	+++	++	+++	++	+++	+++ (27)
Stability and speed	+++	+	-	+++	+	+++	+++	++	-	++	+++	++ (23)
Usability	+++	+	+++	+++	++	+++	+++	++	+++	++	+++	+ (23)
Effectiveness, reliability, and validity	+++	+++	-	+++	++	+++	+++	++	-	++	+++	+++ (28)
Logistical and practical obstacles, feasibility and limitations	+++	+	-	+++	+++	+++	+++	++	-	++	+++	++ (23)
Acceptance	+++	+++	-	+++	+++	+++	+++	++	++	++	+++	+++ (27)
Rank	+++ (18)	++ (12)	- (3)	+++ (18)	++ (13)	+++ (18)	+++ (18)	++ (12)	+ (8)	++ (12)	+++ (18)	

2. Piloting and Evaluation – General Remarks

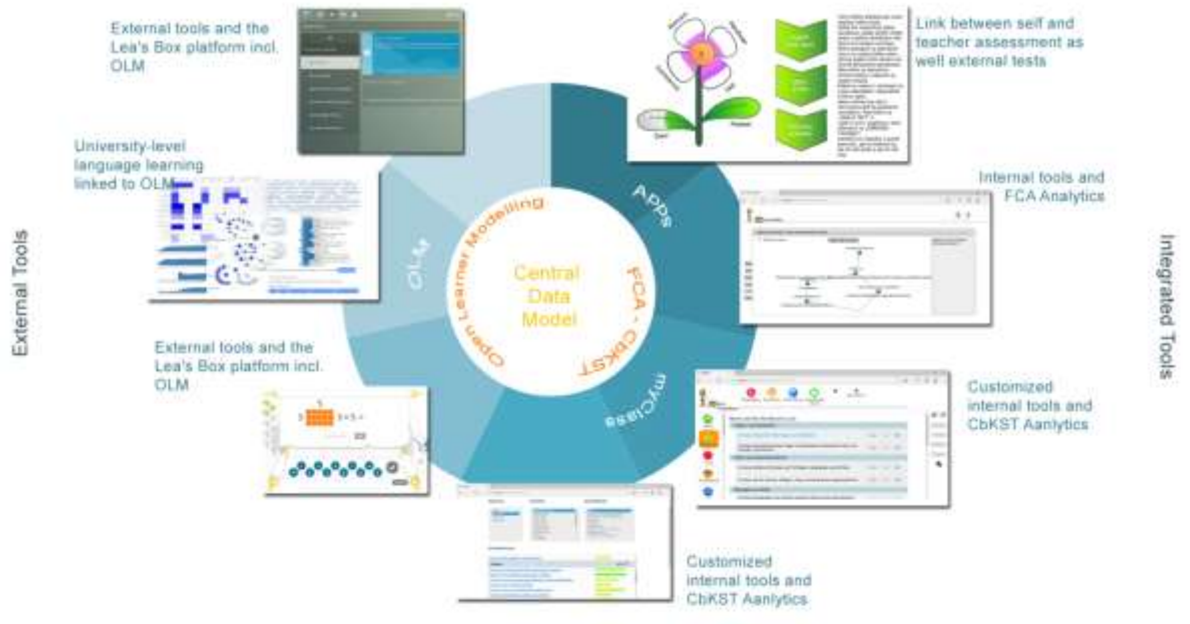
Let us now describe the main approaches used when piloting and evaluating the Lea's Box system.

2.1. Description of the System

In the third and final year of the project, we focused on evaluating the entire Lea's Box system. The system can be described as a web platform for teachers and learners. It has the following properties and functionalities:

- the platform provides links to existing components and interfaces to a broad range of sources for educational data. In such a way, teachers will be able to link the various tools and methods they are already using in their daily practice and which provide software APIs (for example Moodle courses, electronic tests, Google Docs, or other educational tools) together in one central location.
- The platform hosts a set of existing tools and web services to provide an initial set of functions for teachers. These components support activity tracking, domain modelling, and visualization of educational data. The components primarily come from the consortium background of existing developments, tools, and products. The component portfolio is available in system release deliverables D2.4, D2.5 and D2.6.
- The web platform also hosts newly developed LA/EDM services, empowering educators to conduct competence-based analytics of the rich data sets. Modular components are developed to filter, streamline, and aggregate data coming from various sources, to analyse and interpret these data, and to store them in a secure way. Special measures are taken to address data protection and privacy requirements. The set of modules and services as well as data streams are controlled by a superordinate component, the central executive. The concrete properties and designs are available in system release deliverables D2.4-6.
- Furthermore, the web platform provides teachers and learners with existing and newly developed components for visualizing the data and reporting the results of the analyses. Special focus of the research in the project is to develop network and lattice based techniques, such as Hasse diagrams, adapt them to the understanding and expectations of end users, and apply them for user model negotiation.
- Finally, the web platform provides interfaces and links to export/report the data and to transfer them to external tools such as the OLM platform of UoB, ePortfolios, or learning management systems.

The following image illustrates the dimensions of the use cases covered by the final system, ranging from internal applications to integrating external tools, from activity tracking to visualizations and Open Learner Modelling.



Dimensions of Use Cases in Lea's Box.

2.2. Ethical aspects of piloting and evaluation

When piloting the Lea's Box system, we had to bear in mind certain specifics that need to be respected when interacting with real users, as stated in the Annex document. These were:

1. work with real personal data – as mentioned in the Annex document, particular emphasis had to be placed on anonymization (via hashing procedures);
2. availability of teachers and students – it was necessary to bear in mind that the organization of the school year, national holidays and national testing may interfere with piloting and evaluation activities. At times, this led to slight changes in the planned course of piloting sessions;
3. possible disruption of the established ways of teaching – in order not to discourage schools from using the tools in the future, it was important not to place too heavy a burden on teachers and schools. Due to many other activities teachers and schools are required to engage in (by the school management or by the national curriculum), it was necessary to make the piloting sessions as non-invasive and non-disruptive as possible. Therefore, even when testing the system as a whole, we decided to use the use case driven strategy, as described in the Annex. This means we first asked teachers and students about their general opinion on the entire system and then focused on one or two specific components of the system. This strategy proved very effective, because this way, by sharing the feedback on individual components within the consortium, we managed to get in-depth information on all main components of the system without causing any major disruptions of normal classroom practices;

4. impact on students – incorrect data may damage attitudes, motivation and further learning path of a student and the willingness of teachers to use the tools in the future, therefore, as mentioned in the Annex, it was important to ensure the following:
5. informed consent (i.e. all end users were informed of the main aim of the project and the pilot studies they were about to participate in);
6. transparency (i.e. we tried to ensure that all teachers and students understood the main purpose of piloting and were aware of the fact that errors may occur; also, they were explained how and where to report possible errors and inconsistencies found within the system);
7. legitimate purpose (i.e. all end users knew what data were to be collected and what their subsequent use was going to be);
8. non-discrimination (i.e. when interacting with end users, we tried to make it clear that precautions are being taken against all possible forms of discrimination based on gender, social background etc.).

To sum up, we were very much aware of the high level of reliability, data protection and effectiveness expected by the end-users and other stakeholders, such as parents or public administration, when we planned, conducted and evaluated our studies.

In order to be in line with the ethical requirements we increased our collaboration with our project Advisory board where ethical issues were one of the main questions discussed. Furthermore, our internal policies within each partnering organization are quite strict regarding meeting standards, therefore we presume that the challenges raised by points 1-8 have been successfully overcome.

2.3. The general approach to piloting and evaluation

Piloting and evaluation was done using the aforementioned framework which constitutes a strategy focused on individual use cases represented by individual tools in the Lea's Box portal. After launching the portal as a whole, teachers were asked to evaluate the overall functionality, but even then we always placed a special emphasis on one or two tools at a time in order to get more in-depth feedback and minimize the burden placed on teachers.

It must be highlighted that all teacher and school participated voluntarily with the project. In addition, all activities in the context of WP5 occurred in the "real life" of schools, teachers, and students. This setting, ultimately, demands adjusting evaluation and piloting activities to the real needs and ideas of teachers as closely as possible.

The number of pilots and evaluation studies corresponds to the plans, although the time of events sometimes differs from the original plans. Again, the reason is the need to incorporate the organization of the school year as well as real demands and interests into the piloting plan and also a reason is the fact that the original development activities have been delayed at some points.

The main points we focused on during the evaluation phase were functionality and adequacy, stability and speed, usability, effectiveness, reliability and validity, logistical and practical obstacles, feasibility and limitations, and acceptance (based on the general framework).

Dimensions of Use Cases in Lea's Box.

Evaluation variable	Typical evaluation questions
Functionality and adequacy	Do the functions provided answer end-users´ need? Are the functions adequate to the needs? a
Stability and speed	Is the operation flawless?
Usability	Are the tools usable in real-world environment?
Effectiveness, reliability, and validity	Do end-users agree with interpretation of the results?
Logistical and practical obstacles, feasibility and limitations	Are there any obstacles on the way to incorporate the tools into routine operation?

While in years 1 and 2 of the project, the main focus of evaluation and piloting activities was to design the system and its features and to adjust the user interfaces and functionalities according to the demands of educators. The focus of the final period was on supporting the dissemination and exploitation efforts. This changed the concrete realization of the activities in the 3rd period as opposed to the prior periods. The way the system and the features evolved has been reported in the system release documents. We started from the idea of a conventional, rather monolithic, web portal, shaped this to a more modular and app-like style and concluded with a servitization approach where the portal itself has taken a back seat and where the individual features, functionalities, analyses, and services took centre stage.

3. Piloting and evaluation studies in the Czech Republic

The following table summarizes the plan as described by the Annex (left) and the actual activities (right) carried out in year 3 in the Czech Republic. Evaluation activities always directly followed pilot activities.

Piloting	
Flower tool May 2016 – June 2016 50 learners, 3 teachers	Flower Tool May 2016 – June 2016 48 learners, 3 teachers
Entire system pilots September 2016 – November 2016 100 learners, 4 teachers	Entire system pilots September 2016 – December 2016 426 learners, 18 teachers
Evaluation	
Flower tool May 2016 – June 2016 50 learners, 3 teachers	Flower Tool May 2016 – June 2016 47 learners, 3 teachers
Entire system pilots September 2016 – November 2016 100 learners, 4 teachers	Entire system pilots September 2016 – December 2016 70 learners, 11 teachers

3.1. Piloting and evaluation aspects in general: Czech Republic

Bearing in mind the overall goals of the project piloting and evaluation activities were shaped in a close collaboration with training, dissemination and exploitation activities of the project. As this was the final period thereof, we heavily collaborated with our partnering organizations and advisory board consisting of Mr Václav Nádvorník and Mrs. Adéla Zelenda, both senior experts in novel approaches to education with sufficient experience from schools daily routine (Mr Nádvorník) and familiar with the best practice in maintaining ethical standards in work with children and schools (Mrs. Zelenda).

Therefore we aimed carrying out such activities that

a) provide the project with information needed for the successful completion (see evaluation matrix)
b) can directly lead to or facilitate dissemination, exploitation and impact.

All of the activities also followed the recommendations and requirements as set by the Annex of May 2016.

3.2. Piloting and Evaluation: Flower Tool

Deliverable 5.5 introduced the so-called Flower Tool. The unique functionality of the Flower Tool is the fact that it allows users to combine different sources of assessment – self-assessment, teacher assessment and external assessment (i.e. assessment by an established test). This comparison allows teachers and students to identify possible problems and to take appropriate action (e.g. to identify which students are underestimating themselves) in an easy and straightforward way. Another unique feature of the tool is the fact that it includes assessment of several important aspects that may play a very important role in students' learning process but that frequently remain unnoticed (motivation, effort, easiness, satisfaction).

Deliverable 5.5 described the outcomes of the piloting activities carried out in autumn 2015 and the first findings of the Flower Tool evaluation studies. These evaluation studies were conducted via questionnaires and focus groups, for details see deliverable 5.5.

In general, users were satisfied with the tool per se, but complained about some technical aspects and also about the visualization techniques used in the tool. Therefore, we decided to update the tool based on users' suggestions and to complete the piloting activities in spring 2016.

In spring 2016, one of the participants in the piloting studies was ZŠ Londýnská, Prague. This turned out to be very beneficial for the following reasons. Firstly, this school had also participated in the co-design activities and was thus able to give us very detailed and thorough feedback regarding the functionality of the tool (especially the vice-director, Mr. Nádvorník, who is also a member of the advisory board). Secondly, pupils in this school use a very wide variety of ICT components (different operation systems, different hardware, different internet connection set-ups etc.), which enabled us to adapt the tool to a wide range of technical solutions that may be employed at schools. This was achieved thanks to a direct participation of Scio's ICT experts in the piloting activities – they were present at the school during the piloting itself and solved any unexpected issues immediately. Thanks to this, the tool should currently run smoothly regardless of the operation system or hardware used.

Further evaluation was done using the same questionnaire as in autumn 2015 and the findings were quite similar to those described in D5.5.

3.3. Piloting and evaluation of the entire system

Based on the evaluation of the first Flower Tool, we modified and extended the tool and better incorporated it into the whole system. Thus, the second pilot study comprised the whole system with a special focused placed on the updated Flower Tool. Although this tool at first sight appears a disconnected add on to the system, it in fact mirrors the concept of a full system best. At the surface we have a functionally simple tool with a neat user interface. This, however, is only the front end that covers three major external data sources: We have the data source based on teachers' evaluation, the students' self-assessments, as well as the link to the external commercial testing product of SCIO, which is directly and seamlessly linked to the system. This 'surface' too is directly connected to the main system through Lea's API. Teachers can access the data with the portal and add further data sources or make further assessments. Of course, the data analyses can be accessed in the system (e.g., the OLM or the CbKST analysis). Insofar, this flower tool is the prime prototypical scenario for a "full system use".

The main updates of the Flower Tool included the following changes:

- Teachers said they would like to use the tool in other subjects as well, so we extended the functionality to other domains (Czech, Maths, General Academic Prerequisites, Learner Autonomy); please note these domains also include skills that are not so commonly evaluated, such as critical thinking (which is part of the General Academic Prerequisites test) and Learner Autonomy, i.e. the ability to learn autonomously on one's own (this concept includes the ability to set meaningful goals, to choose appropriate ways to reach a goal, to be able to evaluate whether a goal has been reached and also to perceive mistakes not as a failure but as a source of new experience);
- Teachers and students had some objections with respect to the design of the tool so we modified it in order to make it look nicer and more user-friendly;
- Teachers had troubles understanding the results generated by the tool, so we simplified and changed the way the results are displayed – we replaced complicated graphs with simpler, straightforward flowery design which allows users easily compare test results with self-assessment and teacher assessment (see an example below).

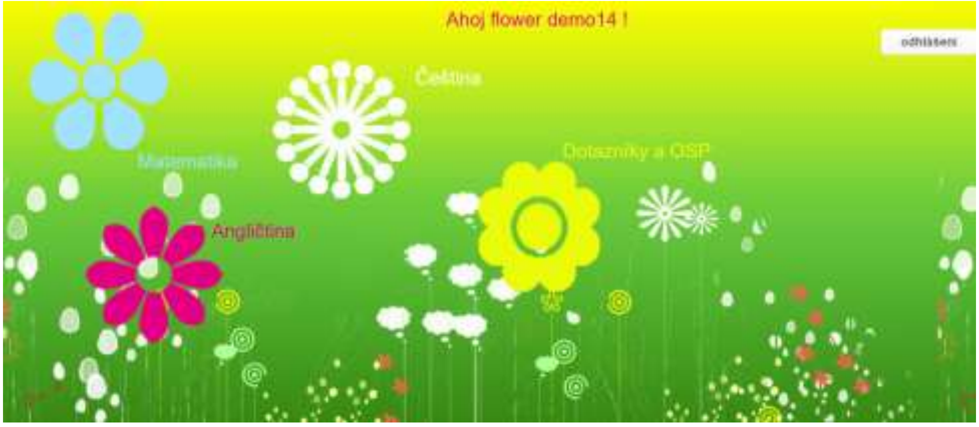
3.3.1 Different evaluation scenarios

The updated Flower Tool embedded in the Lea's Box system was then piloted in several different types of school and educational institutions in order to determine how well the system works in different educational contexts. Therefore we defined 4 different evaluation scenarios for typical different end users with different needs. The aim of each evaluation was to identify where the system can be utilized in different environments. These evaluation scenarios were based on a setting of:

1) A large public school
2) A private school
3) An orphanage
4) A small village school (with mixed-age classes)

The updated tool now looks as follows:

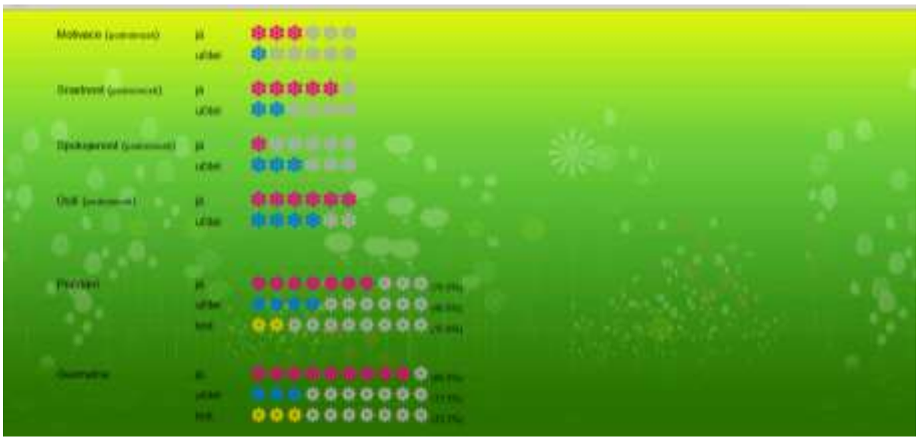
1) Main screen: each flower represents one subject (Maths, Czech, English and General Academic Prerequisites, which also includes the Learner Autonomy questionnaire).



2) By clicking on one of the flowers, the following screen appears. Just like before, the student first evaluates him/herself in several subject-related skills (each skill is represented by one leaf, in Maths – displayed below – the skills are Numerical literacy and Geometry) and in four aspects related to the learning process (Motivation, Easiness, Effort, Satisfaction), each aspect being represented by one petal.

3) The self-evaluation is done in the form of a simple questionnaire. We strived for a maximum simplicity and straightforwardness of the questions included, because teachers had pointed out that in the original flower, some of the questions were too complicated for the students to understand. An example of a questionnaire is depicted below. Depending on the answers, the corresponding proportion of the leaf/petal is coloured in. Students also need to write what grade they believe they deserve into the middle of the flower.

4) After students have completed the self-assessment questionnaires, they are redirected to one of the tests. These had been developed by Scio earlier (i.e. not within this project).
5) Teachers evaluate each student separately. They do so in a table, i.e. not using the flowers, which saves their time and makes the evaluation more user-friendly.
6) Finally, students and teachers can compare self-assessment, teacher assessment and external assessment (test results). This applies to the subject-related skills; for grade, motivation, effort, satisfaction and easiness, only a comparison of self-assessment with teacher assessment is available.



7) Apart from the individual results, teachers can also display summarized results for the entire classroom.

It is important to point out that the Flower Tool serves only as the first step that needs to be followed by further action, i.e. the usefulness of the tool depends mainly on the way teachers are going to use the results obtained via the tool. There may, of course, initially be some problems caused by misunderstanding of the tool's functionality or by a lack of relevant experience. For example, kids may fill in the self-assessment questionnaires incorrectly in order to make the flower look as colorful as possible or there may be substantial discrepancies between teacher assessment, self-assessment and test assessment. The solution is, however, fairly straightforward – teachers need to discuss the results with their students, repeatedly clarify the purpose of self-assessment, think about possible explanations of the discrepancies and take appropriate action. In other words, they need to perceive the Flower Tool as a source of information, but not as an immediate solution. Without these subsequent steps, the tool itself has no purpose.

Currently, not all teachers in the Czech Republic are able and willing to take such steps though. Czech assessment culture still bears marks of the communist era, and self-assessment, evaluation of soft-skills etc. have begun to find their way into the classrooms only recently. However, teachers have been increasingly calling for changes in the assessment culture and have shown increased interest in workshops and courses related to this matter. As changing the assessment culture is a necessary prerequisite to the successful introduction of the Lea's Box tools into the classrooms, we have been organizing extensive workshops, training and dissemination activities with the aim to make Czech assessment culture more progressive and pupil-oriented. These activities are described in the training and dissemination deliverable (D5.8).

3.3.2 OUTCOMES

Let us now focus on the evaluation of the system. Lea's Box was initially evaluated using an online questionnaire[1], which provided us with a brief overview of its general suitability. To see how well the system suits different learning contexts (state schools, private schools, small village schools or orphanages), we conducted several follow-up studies. These were done in the form of a focus groups

or video-conferences. More focus groups are to follow in January 2017 as a part of our exploitation activities. Even though the project will have finished before that, we, as a business member of the consortium, have concluded that other schools or different people within the same schools might be interested in using the tool. Therefore, we are going to engage in direct exploitation activities as soon as possible. This conclusion is based on the positive feedback we received during the piloting and evaluation sessions described in this deliverable.

Photo from a CZ 2016 pilot.

[1] Questionnaire design was based on methodology described in *Iarossi, G., 2006. The power of survey design: A user's guide for managing surveys, interpreting results, and influencing respondents. World Bank Publications*

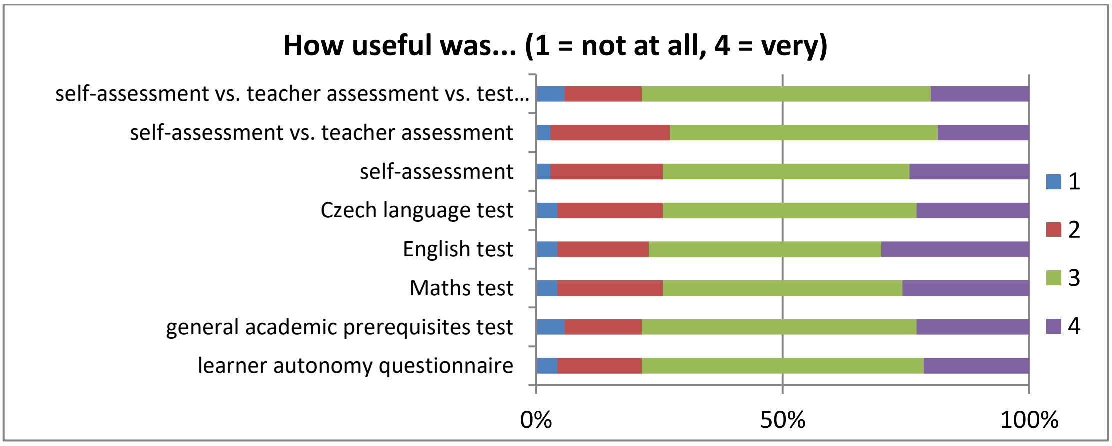
Let us now summarize the main findings of the evaluation studies. The first step of the evaluation was, as mentioned above, an online questionnaire, which was designed in such a way so as to address the issues outlined in the general evaluation matrix (see the introduction). The following table contains a summary of questions in the student questionnaire and teacher questionnaire (letters S and T indicate, whether that particular question was included in the student questionnaire, teacher questionnaire or both) and their relation to the general evaluation matrix. If not indicated otherwise, we used a 4-point Likert scale. The questions in the student questionnaire were formulated in such a way so as to be easily understandable for the target age-group, because it has been shown that age-appropriate formulations increase the accurateness of the answers[2]. The questionnaire was filled in by 70 students and 6 teachers.

Functionality and adequacy
How useful was the learner autonomy questionnaire? (S, T)
How useful was the general academic prerequisites test? (S, T)
How useful was the English test? (S, T)
How useful was the Math test? (S, T)
How useful was the Czech language test? (S, T)
How useful was self-assessment? (S, T)
How useful was the comparison of self-assessment and teacher assessment? (S, T)
How useful was the comparison of self-assessment, teacher assessment and test assessment? (S, T)
Stability and speed
How often did the system crash? a) never b) almost never c) sometimes d) often e) all the time (S, T)
The system reminded you of a a) starving cheetah b) freaked out hen c) athletic snail d) half-dead turtle (S)
Did you experience any delays when using the system? (T)
Usability
Should your teacher use the results obtained via the system? If so, how? If not, why not? (S)
Would you like to use the results obtained via the system? a) yes and I know how I'd do that b) maybe and I have a few ideas how I could do that c) no, I don't understand them d) no, I don't find them useful (T)
How well do you understand the results presented by the system? (S, T)
Were the results displayed in an understandable way? (S, T)
Effectiveness, reliability and validity
Did the results displayed by the system meet your expectations? (S, T)
Do you trust the results of the learner autonomy questionnaire? (S, T)
Do you trust the results of the general academic prerequisites test? (S, T)
Do you trust the results of the English test? (S, T)
Do you trust the results of the Math test? (S, T)
Do you trust the results of the Czech language test? (S, T)
Do you trust students' self-assessment? (T)
Logistical and practical obstacles, feasibility and limitations

[2] Copple, Carol, and Sue Bredekamp. Developmentally appropriate practice in early childhood programs serving children from birth through age 8. National Association for the Education of Young Children. 1313 L Street NW Suite 500, Washington, DC 22205-4101, 2009.

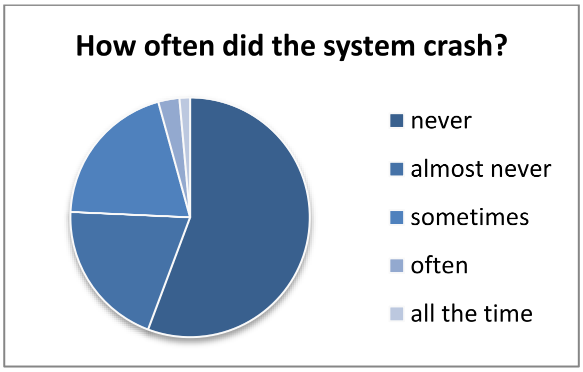
If you decided to use the system on a regular basis, which of the following obstacles would you have to deal with? (T)
- Obstacles caused by the management of the school (Y/N)
- Technological obstacles – lack of PCs, laptops (Y/N)
- Technological obstacles – internet connection (Y/N)
- Time (Y/N)
- Other (Y/N)
Acceptance
Would you recommend the system to students from a different classroom? (S)
Would you recommend the system to your colleagues? (T)
Would you like to use the system on a regular basis? (S, T)
Would you recommend your teacher to buy the Flower Tool? (S)
What would you prefer: the Flower Tool or the tests only? (S, T)
What do you think your teacher would prefer: the Flower Tool or the tests only? (S)
What do you think your students would prefer: the Flower Tool or the tests only? (T)
Would you like to use the Flower Tool in other subjects? (S, T)
Other
Do you have any suggestion regarding the Flower Tool? (S, T)
Do you have any suggestions regarding the design of the Tool? (T)

Let us now summarize the results. As for the functionality and adequacy, the table below shows that most children (about 80 %) perceive the functions offered by the tool (tests, self-assessment, its comparison with teacher assessment etc.) as useful or even very useful.

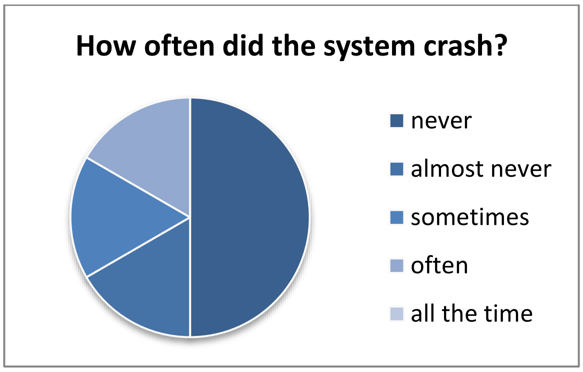


With the exception of the English test, 4 out of 6 teachers who filled in the questionnaire perceived the tests as very useful (i.e. they chose number 4 on the Likert scale); one teacher chose number 3 and one teacher chose number 2. 50 % of teachers perceived self-assessment as quite useful and 50 % as very useful. Similarly, a third of the teachers perceived the comparison of self-assessment and teacher assessment as quite useful and two thirds as very useful; the same applies to the comparison of self-assessment, teacher assessment and external (test) assessment.

As for the stability and speed, 56 % percent of students said the system never crashed and only 3 % of students said it crashed often or all the time. Similarly, 50 % of teachers said that the system never crashed and nobody said it kept crashing all the time. There may be a variety of reasons behind occasional failures of the system – apart from errors in the system itself namely problems with internet connection. About 70 % of the students said the speed of the system reminded them of a starving cheetah or a freaked out hen. Only 9 % of the students said it reminded them of a half-dead turtle. As for the teachers, 5 out of 6 teachers experienced no or only minor delays when using the system, while only 1 teacher experienced severe delays.

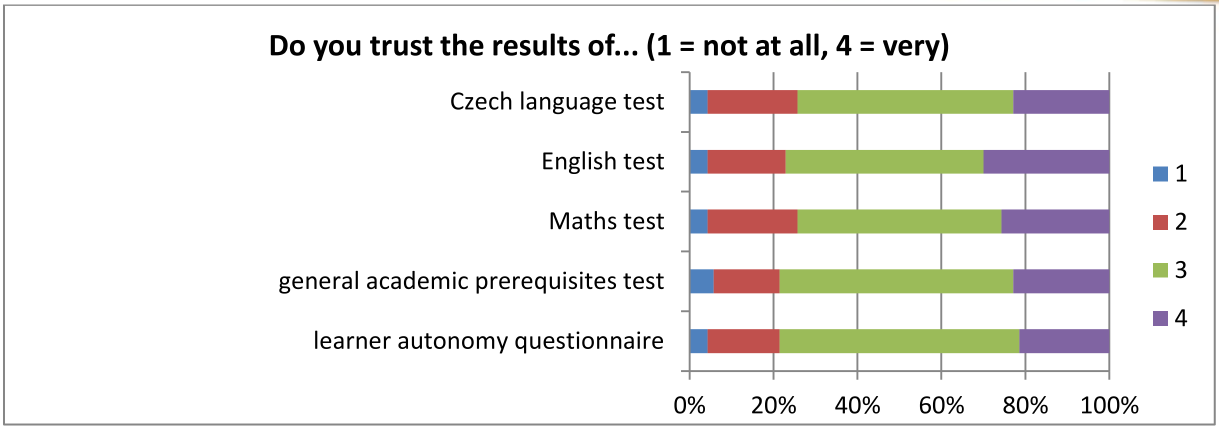


students

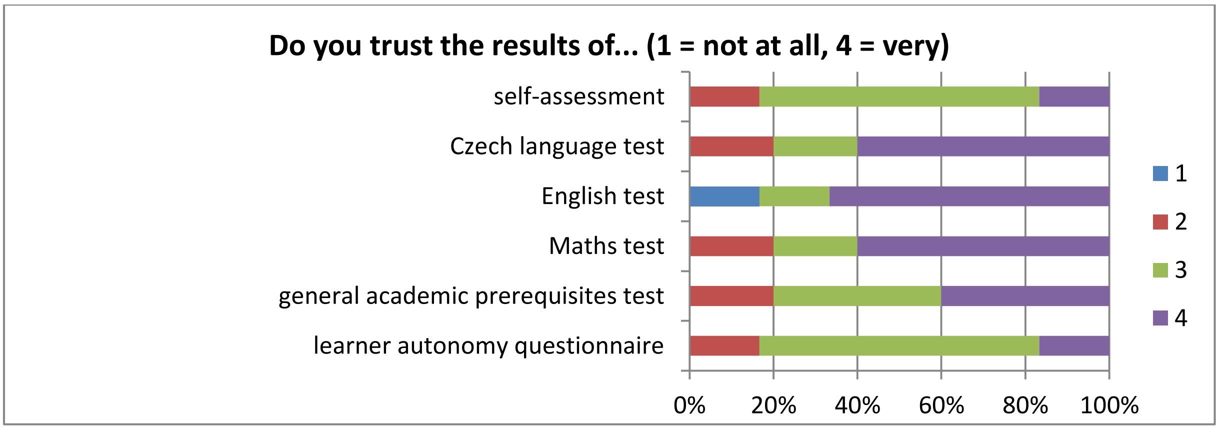


teachers

As for the usability, 80 % of the students thought their teacher should somehow use the results obtained via the system. Only 1 out ot 6 teachers said they wouldn't use the results obtained via the system any further. 84 % of the students said they understood the results displayed by the system quite well (50 %) or even very well (34 %). Most teachers also said they understood the results well (17 %) or even very well (50 %), but two teachers (i.e. 33 %) said they did not understand them much. About 80 % of the students and 83 % of the teachers said the results were displayed in quite an understandable (students 41 %, teachers 50 %), or even very understandable (students 37 %, teachers 33 %) manner. As for the effectiveness, reliability and validity, 74 % of the students said the results presented by the system met their expectations (57 % percent chose number 3 on the Likert scale while 17 % chose number 4). In the same question, five out of six teachers chose number 3 on the Likert scale while only one teacher chose number 2. About 80 % percent of the students said they quite trusted or even very much trusted the results displayed by the system, see table below. Teachers' opinion was quite similar, see figure below.



students



teachers

As for logistical obstacles, 1 out of 6 teachers said internet connection would be an obstacle if they decided to use the system, and 5 out of 6 teachers said time would be an obstacle. Only 1 teacher said there would be no obstacles to using the system whatsoever.

Finally, as for the acceptance, the results were as follows:

- 37 % percent of students would quite recommend the system to students from a different classroom and 44 % would highly recommend it, while 33 % of teachers would probably not recommend the system to their colleagues and 67 % would quite recommend it;
- most students would quite like (39 %) or very much like (20 %) to use the system on a regular basis; 83 % of the teachers said they would quite like to use the system on a regular basis;
- 72 % of the students would recommend their teacher to buy the Flower Tool;
- 80 % of the students would prefer the Flower Tool to just taking the tests it contains (however, interestingly, only two thirds of the students think their teacher would choose this option as well);

- in comparison, half of the teachers would prefer the Flower Tool to just making students take the tests (however, 67 % of the teachers believe students themselves would prefer the Flower Tool);
- 30 % of the students would quite like to use the Flower Tool in other subjects as well and 34 % of the students would very much like that; in comparison, 67 % of teachers said they would quite like to use the Flower Tool in other subjects but the rest said they wouldn't like to do that.

The second part of the evaluation was focus groups and phone-call sessions with the representatives of the schools and institutions where the piloting had taken place. We used structured interviews to get more context-specific and in-depth feedback related to the suitability of the system in that particular educational context, with a special focus placed on the areas described in the general evaluation matrix. The main findings were as follows:

Large public school:

- As for the functionality of the system, teachers liked the fact it included self-assessment and the evaluation of subject-related behaviour and attitudes (motivation, learner autonomy etc.). As for teacher assessment, teachers were at times uncertain about how to assess pupils in domains such as motivation. More training and experience would be needed to make them more experienced in this type of assessment.
- Teachers said the average time pupils spent using the system before their results were displayed was adequate and acceptable. However, they said the stability and speed could still be improved. The system was, at times, quite slow, which was probably caused by slow wireless connection combined with high demands on the internet connection placed by the system.
- The results presented by the system seemed to be valid and teachers trusted them, with the exception of occasional errors in the database, in which case the system displayed erroneous results. Nevertheless, this issue has already been solved.
- When asked whether they would prefer the Lea's Box system to what they normally use, all teachers unanimously opted for the Lea's Box system. As the school where the evaluation took place is one of our established partners, we are preparing several follow-up activities we could offer to this school as a part of the exploitation of the system.

Private school:

- As for the functionality of the system, teachers especially appreciated the possibility to combine different source of information, such as self-assessment and teacher assessment.
- The stability and speed of the system was good, the system did not crash, the only problem teachers mentioned was the fact that some pupils had to log in repeatedly. Teachers said the system could be easily used on a regular basis without any major obstacles.

- Teachers doubted the results presented by the system a bit, because the tasks and evaluation materials presented by the system were rather lengthy and demanding and they believed some pupils may have lost focus.
- They said they could imagine using the system on a regular basis if the issues mentioned above were solved.

Orphanage:

- As for the functionality of the system, teachers said it was satisfactory and did not come up with any specific functionality that was missing.
- The stability and speed was generally good with some exceptions. They didn't mention any obstacles to using the system.
- Overall, teachers trusted the results presented by the system with the exception of disadvantaged pupils or pupils with disabilities, where they were a bit doubtful. They said that for these pupils, the system might be too complicated.
- When asked whether they would like to keep using the system, they said they would indeed like to do that provided it was free.

Small village school with mixed-age classes:

- Among all the functionality offered by the system, teachers above all praised the fact it offers tools for self-assessment, because so far, there is no other tool of such kind.
- The stability and speed was good, the system did not crash. If the system should be used on a continuous basis, there would be no logistical obstacles to using the system with the exception of lack of time.
- Teachers said they trusted the results presented by the system. They also said it unfortunately confirmed the fact their pupils struggle with reading.
- When asked whether they would like to keep using the LEAs Box system, they said that probably yes (with 60% probability), but they would like to improve visualization options.

As a complement of these evaluation studies, we also ran studies focused on evaluating pedagogical problems (see table below), which we integrated into training and dissemination activities. For each question, we used a Likert scale to determine, to which degree the Lea's Box tools would be a solution to that particular pedagogical question. The table below displays the answer that appeared most frequently (i.e. modal answer).

PEDAGOGICAL PROBLEM	Lea's Box WOULD BE A SOLUTION
1. How to create accurate personalized assessment relevant to each student's progress of learning content?	AGREE
2. How to assess progress? How to assess value added by school/teacher?	STRONGLY AGREE

3. How can students/teachers meaningfully reflect on student work to create higher quality work?	STRONGLY AGREE
4. How does information disseminated by peers as well as peer-to-peer learning effect knowledge acquisition?	DISAGREE
5. How do assessments promote student understanding and growth?	AGREE
6. Which assessments measure best?	AGREE
7. How to find an ideal balance between the urge to gather assessment data and having sufficient time for teaching? What can be omitted and what definitely needs to be recorded?	NEUTRAL
8. How to vary instructional strategies to increase learning efficiency?	NEUTRAL
9. What assessment alternatives can be incorporated in order not to focus solely on grades?	NEUTRAL
10. How can efficiency of new teaching/learning models be assessed?	NEUTRAL
11. How can teachers save time while creating individualized development/learning plans?	AGREE
12. Who are the best/good and who are the low performers?	AGREE
13. What to teach next - /wrt to the entire class and /wrt to individual students?	AGREE
14. How to build groups among students (e.g., for projects, presentations, homework, etc.) – which students match which don't?	AGREE
15. What materials / topics should individual students attend – what homework would be the most effective for an individual?	STRONGLY AGREE
16. What do have students in common, where are they different?	AGREE
17. Which social standing do students have in a group?	NEUTRAL
18. What test items and evaluation task would be optimal for a particular student?	NEUTRAL
19. What are suitable learning paths?	AGREE
20. How is a domain structured?	NEUTRAL
21. What are the differences between students' perception of a domain in comparison to the teachers'?	NEUTRAL
22. How to evaluate one's own teaching or specific educational projects?	AGREE
23. Is the teaching material appropriate?	AGREE
24. Do students reach the goal of the course?	AGREE
25. Is there a risk for particular students to fail in a course?	NEUTRAL
26. How to address cross-subject, cross-domain meta-skills within a class or across the entire school?	AGREE
27. How to give tailored, understandable, and formative feedback – students – to parents – how to inform colleagues?	STRONGLY AGREE
28. How to select suitable learning materials / modalities for individual students (in terms of learning styles, for example)?	NEUTRAL
29. How to profit from all the various technologies that are available (apps, games, online tools, etc.) best?	AGREE

3.4. Evaluation of the specific business cases in the Czech Republic

In addition to the above mentioned case, Scio also researched some of the business cases relevant for further exploitation of the Lea´s Box outputs.

#SCIO_EvalStudy of Business Case 1: Value-Added Feature in a SCIO product:

This study was carried out with Mozartova elementary school in Olomouc. This school is SCIO´s VIP client with extensive knowledge and experience of the existing (not only Scio´s) assessment and analytical products. The unanimous verdict was that Lea´s Flower app would be a massive improvement to the reporting possibilities of the currently available tools on the market and would be a reason to buy.

#SEBIT_EvalStudy of Business Case 5: Learning Analytics for HEIs

This business case was discussed with Marian Golis, who is our senior expert in designing analytical tools for HE (e.g. formerly responsible for entry examinations to approximately 1/3 of all Czech universities) . He told us that currently the possibilities of a massive uptake of learning analytics in higher education cannot be expected due to the persisting cultural and personal situation of many Czech universities. However, there is a law being drafted obliging universities and other higher education institutions to increase their activities in the area of evaluation. According to his words, there may a market niche for some of LB tools. This possibility will be further researched after the law is implemented.

4. Piloting and evaluation activities in Austria

In Austria (and Germany) we pursued the following activities, summarized by following table. First we continued and completed the so-called myClass pilots with the partner schools in Austria and Germany. In addition we established an intensive cooperation with a partner school in Graz and evaluated various system components. Along with the various events with teachers we also present an evaluation of the foundations of the entire system and its features, that is, the pedagogical questions. Finally, we present a simulation study focusing on the LPV, the performance prediction component.

Piloting	
myClass / portal / full system March 2016 – June 2016 400 learners, 50 teachers	Flower Tool, FCA Tool March 2016 – December 2016 34 learners, 1 teacher
Evaluation	
Full system / Pedagogical Questions Nov – Dec. 2016 50 teachers	LPV - Simulations / ext. data sets

4.1. MYCLASS PILOTS IN AUSTRIA AND GERMANY

As reported in previous reports, over the entire project duration we had 2 pilot studies on the basis of the myClass tool. This too originally was developed in the context of the Next-Tel project (www.next-tell.eu). In the course of the Lea's Box project we substantially revised the tool and changed the look and feel.

Taus (Gemeinschaftschule in der Taus) is a general school covering primary and secondary levels. The nature of the school is a "collective" one, bringing together students from the 1st to the 10th grade. In total this school has about 650 students and a faculty of about 50 teachers. The school is located in

Backnang, Germany, a small town near Stuttgart. KPH (Praxisvolksschule der Kirchlich Pädagogischen Hochschule Graz) is the training school of the Christian teacher training academy in Graz, Austria. The school has a general catholic spirit and it offers open educational forms such as Montessori or Jenaplan education. The school covers the primary level with 8 classes and about 200 students. There is a faculty of about 30 teachers.

Both schools Taus, and KPH can be characterized as typical technology lean European school. The daily routines are primarily non-digital, even administrative tasks such as making records about not attended school hours, etc. The use of modern electronic media in classroom teaching is mostly limited to presenting contents with a projector. Activities we can summarize include projects for learning how to use Android tablets and scattered use of various learning apps (such as TUG's 1x1 Ninja) . On the secondary level (where there is already informatics classes) in addition we can add 'internet safety' workshops and an teaching Microsoft Word, Powerpoint and Excel. A special situation arises at KPH for the Montessori classes. In Montessori pedagogy are carful and detailed record keeping of all learning activities is required. Usually teachers and students get together at the end of a (usually very free and unstructured) school hour and discuss the activities while the teacher writes down the activities of the students. The aim is to get a systematic overview of thee individual students' activities and learning progress.

As reported in the previous deliverables D5.4 and D5.5, we started from the original myClass tool and substantially revised the tool and its functions in the focus groups of year 1. The original myClass is a larger tool that, in principle, is an electronic class book. It features tools for activity tracking, activity archiving, portfolios, the tracking of learning progress and a set of reporting functions.

Screenshot of the original myClass version.

As reported in the system release documents (D2.4, D2.5, D2.6) for both schools we setup tailored solutions, both based on the same functional modules. Both tools, however, focussed on the tracking of competence-oriented learning performance and on providing related reports. The Taus version (cf. next image) has a special feature to automatically generate and print report cards and booklets).

Screenshot of the myClass version of the Taus school.

Over the entire project duration we accompanied the use of the tools and collected feedback and suggestions on the basis of frequent meetings. A major part of the interaction and work was maturing the prototype software and the bug fixing. Noteworthy is also the need to provide a a “near product” configuration and administration tool. In other terms, over the project duration we advanced both myClass versions to a quality (or readiness) level that is close to that of commercial products. As one major outcome of the project, we provide both myClass versions free for download through the website on an open source basis. During summer term 2016 we introduced and extended the myClass reporting features for the KPH school with the CbKST analytics functions and the Hasse diagram tool as front end. The idea of this approach was found interesting and the tool was used, specifically in the Montessori classes as one means of negotiating activities and learning performance with the students themselves (even these being primary school children at an age range of 6 to 10 years). A more detailed feedback or specific data were not recorded since the feature wasn’t use long and thorough enough.

Screenshot of the myClass version of KPH showing the Hasse diagram visualization.

Both pilot studies have been concluded and finished at the end of the summer term 2016 (in July 2016). This was because summer term 2016 was the last full school term in the lifetime of the project. We had concluding talks with both school’s contact persons (i.e., Julian Dupont for Taus and Erika Wolfberger for KPH) where we collected final feedback and ideas. Both schools expressed their will and intention to continue using the tools on a regular basis as one of their “normal” software tools. As part of the exploitation beyond the project, we will maintain lose cooperation and TUG will provide software support (if necessary) beyond Lea’s Box.

MAJOR CONCLUSIONS:

- A simple tool that is building upon the concept of competencies (and therefore on teaching and pedagogy) is highly appreciated by teachers. The vast majority of software solutions targeting on schools are mere administration tools (e.g., for setting up timetable or keeping records about unattended hours). In Austria we have two solutions that are widely promoted by the school authorities, that is *Untis* (http://www.untis.at/HTML/start.php) and Vizyon (http://www.netop.com/de/klassenraum-management-software/produkte/netop-vision.htm). Austrian teachers (also outside this particular pilot school) expressed high interest in developing "pedagogical" modules for these software solutions to give them an added value for teachers. This clearly has become a part of the project's exploitation strategy.
- As a roll-out strategy we learned that starting with a very simple tool is the best solution for the average teacher. The myClass versions appear a perfect solution. Once teachers have familiarized themselves with such tool and its pedagogical added value, then, slowly and step by step, further features can be introduced, perhaps on a semester or even yearly basis.
- To adopt learning analytics as such, a lengthy processes of familiarizing with ideas, methods, and solutions is necessary. A majority of teachers prefers very simple types of reports and are not interested or even reluctant in deeper or more complex analytics. There is a clear indication for a change in the teacher training across Europe. Data literacy must be promoted much stronger.

4.2. THE CATALOGUE OF PEDAGOGICAL QUESTIONS

OVERVIEW

A major strategy of the Lea's Box project is to drive theory-based Learning Analytics, which we consider superior to purely statistical approach. Such top-down approach comes, for example, with the Competence-based Knowledge Space Theory (CbKST), on which the projects capitalizes widely. Already at the beginning of the project, WP5 collected requirements by teachers. We tried to formalize these as pedagogical questions – questions teachers may ask and questions for which Learning Analytics might have the right answers. In the course of the project, this set of pedagogical questions was a significant starting point for the design and development of features.

Thus, in the context of the evaluation and validation of the entire Lea's Box system, it is critical to evaluate the validity of these pedagogical questions. This means we have to ask whether the catalogue meets the daily practice of teachers, is compatible with the work processes, and it must address the existing gap in the teacher's understanding of learning processes. In this sense of evaluating what has been developed, the completeness of the catalogue is interesting, however, not elementary. A first major aim of this study is to evaluate the validity of the catalogue for Learning Analytics questions. This catalogue, in the end, might also be a valuable asset for future research and development in this area.

MATERIALS

We used the following catalogue of pedagogical questions. We asked participates to rate the relevance / importance of the individual questions for their daily practice as teachers. The rating was made on a scale from 1 to 10 where 10 was the highest rating.

Table. Catalogue of pedagogical questions.

1. How to create accurate personalized assessment relevant to each student's progress of learning content?
2. How to assess progress? How to assess value added by school/teacher?
3. How can students/teachers meaningfully reflect on student work to create higher quality work?
4. How does information disseminated by peers as well as peer-to-peer learning effect knowledge acquisition?
5. How do assessments promote student understanding and growth?
6. Which assessments measure best?
7. How to find an ideal balance between the urge to gather assessment data and having sufficient time for teaching. What can be omitted and what definitely needs to be recorded?
8. How to vary instructional strategies to increase learning efficiency?
9. What assessment alternatives can be incorporated in order not to focus solely on grades?
10. How can efficiency of new teaching/learning models be assessed?
11. How can teachers save time while creating individualized development/learning plans?
12. Who are the best/good and who are the low performers?
13. What to teach next - /wrt to the entire class and /wrt to individual students?
14. How to build groups among students (e.g., for projects, presentations, homework, etc.) – which students match which don't?
15. What materials / topics should individual students attend – what homework would be the most effective for an individual?
16. What do have students in common, where are they different?
17. Which social standing do students have in a group?
18. What test items and evaluation task would be optimal for a particular student?
19. What are suitable learning paths?
20. How is a domain structured?
21. What are the differences between students' perception of a domain in comparison to the teachers'?
22. How to evaluate one's own teaching or specific educational projects?
23. Is the teaching material appropriate?

24.	Do students reach the goal of the course?
25.	Is there a risk for particular students to fail in a course?
26.	How to address cross-subject, cross-domain meta-skills within a class or across the entire school?
27.	How to give tailored, understandable, and formative feedback – students – to parents – how to inform colleagues?
28.	How to select suitable learning materials / modalities for individual students (in terms of learning styles, for example)?
29.	How to profit from all the various technologies that are available (apps, games, online tools, etc.) best?

SAMPLE

Paper-pencil survey: A paper-pencil version of the survey was issued to the attendees of the 3 teacher training events in Austria. In total 17 teachers filled in the survey from the event in Lower Austria in November 2016, 25 from the seminar in Vienna on December 5 and 4 from the seminar in Graz on December 15. In total 46 teachers filled in the survey, the minimum age was 27 the maximum age 53; 27 of them were female and 19 male. The primary school type of teachers was NMS (new middle school) for 5 to 8 graders. This is because this type is rather new and has more resources for realizing stronger / alternative pedagogical strategies in the daily practice of schools.

Online survey: In addition to the training events in November we made a call through the Austrian eLSA network, the network for technology in schools. Only 14 complete responses could be collected, within 4 weeks. The minimum age was 24, the maximum age 43; 9 participants were male, 5 female.

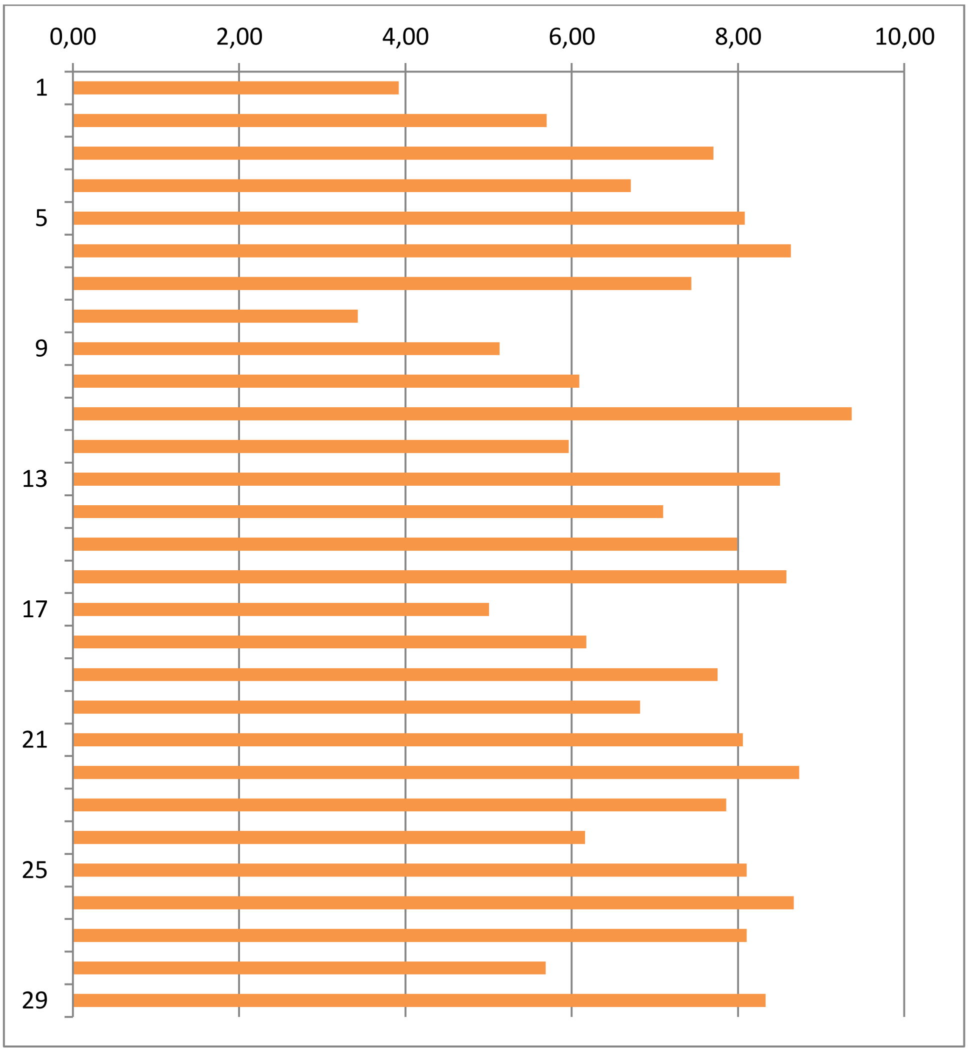
RESULTS

The following table summarize the results. The first column identifies the pedagogical question as of the table above. “Relevance Paper” lists the results of the paper/pencil version (n=46), “Relevance Online” lists the results of the online version (n=14), and “Relevance Total” shows the average relevance rating of both groups (n=60). “Lea’s Solution”, in turn, shows the participants’ judgements of the quality to which the solutions of individual Lea’s Box tools and features address the pedagogical questions and provide answers to them. Equality to the relevance rating, the rating was made on a scale from 1 to 10 (with a maximum of 10). Please note that there is a trade-off between the relevance of a question and the quality, appropriateness, or completeness of the solution/ answer; we invested more research and development efforts for features we and the participants of prior design study and focus groups believed to be of higher relevance.

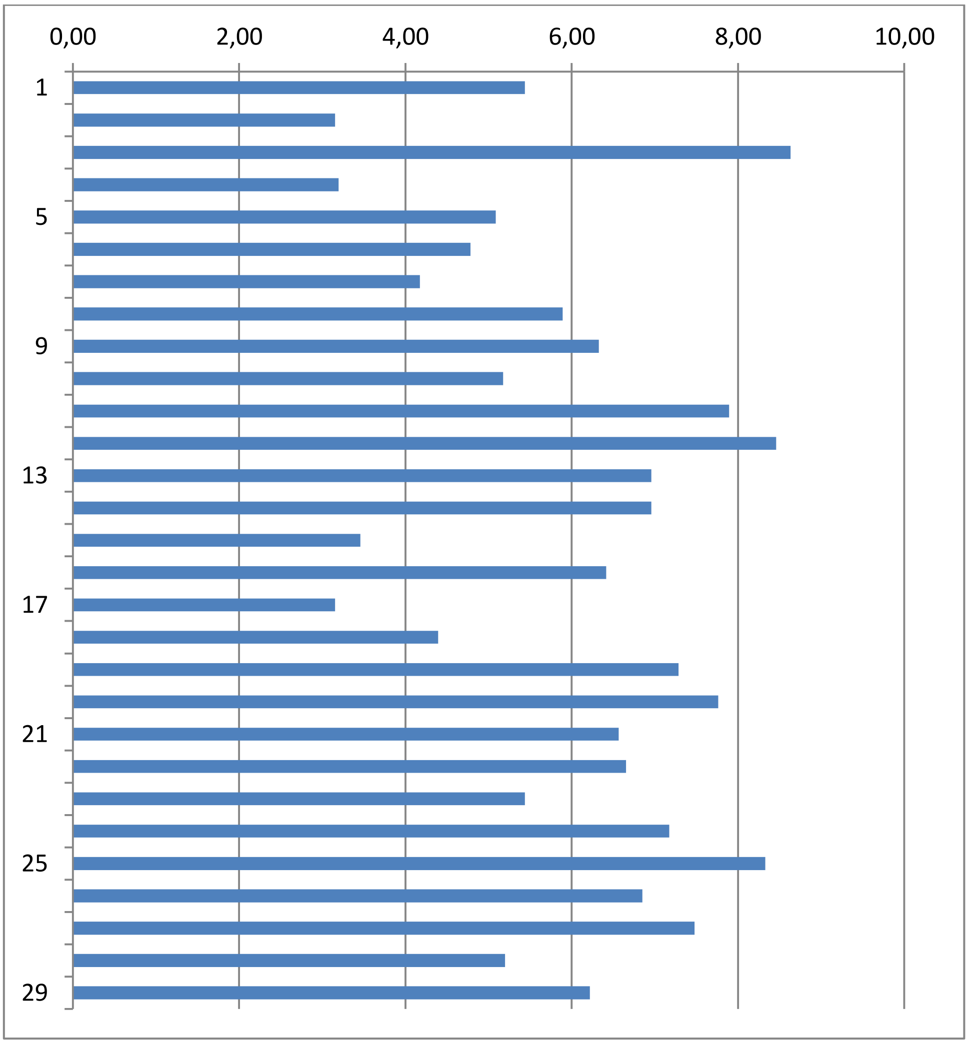
Table. Summary of results.

Question	Relevance Paper	Relevance Online	Relevance Total	Lea's Solutions	Index
1	3,91	3,93	3,92	5,43	21,27
2	6,11	5,29	5,70	3,15	19,26
3	7,76	7,64	7,70	8,63	66,98
4	7,07	6,36	6,71	3,20	22,58
5	8,30	7,86	8,08	5,09	42,24
6	8,63	8,64	8,64	4,78	41,28
7	7,52	7,36	7,44	4,17	31,40
8	3,28	3,57	3,43	5,89	19,34
9	4,98	5,29	5,13	6,33	31,49
10	6,11	6,07	6,09	5,17	31,61
11	9,30	9,43	9,37	7,89	73,42
12	5,78	6,14	5,96	8,46	48,90
13	8,43	8,57	8,50	6,96	58,68
14	7,13	7,07	7,10	6,96	49,60
15	8,26	7,71	7,99	3,46	28,55
16	8,52	8,64	8,58	6,41	54,65
17	4,87	5,14	5,01	3,15	15,35
18	6,07	6,29	6,18	4,39	26,63
19	8,15	7,36	7,75	7,28	59,37
20	6,72	6,93	6,82	7,76	52,13
21	8,04	8,07	8,06	6,57	52,81
22	8,83	8,64	8,73	6,65	58,71
23	7,72	8,00	7,86	5,43	41,94
24	6,17	6,14	6,16	7,17	44,29
25	8,07	8,14	8,10	8,33	67,15
26	8,70	8,64	8,67	6,85	59,55
27	8,07	8,14	8,10	7,48	60,31
28	5,80	5,57	5,69	5,20	30,16
29	8,02	8,64	8,33	6,22	49,87

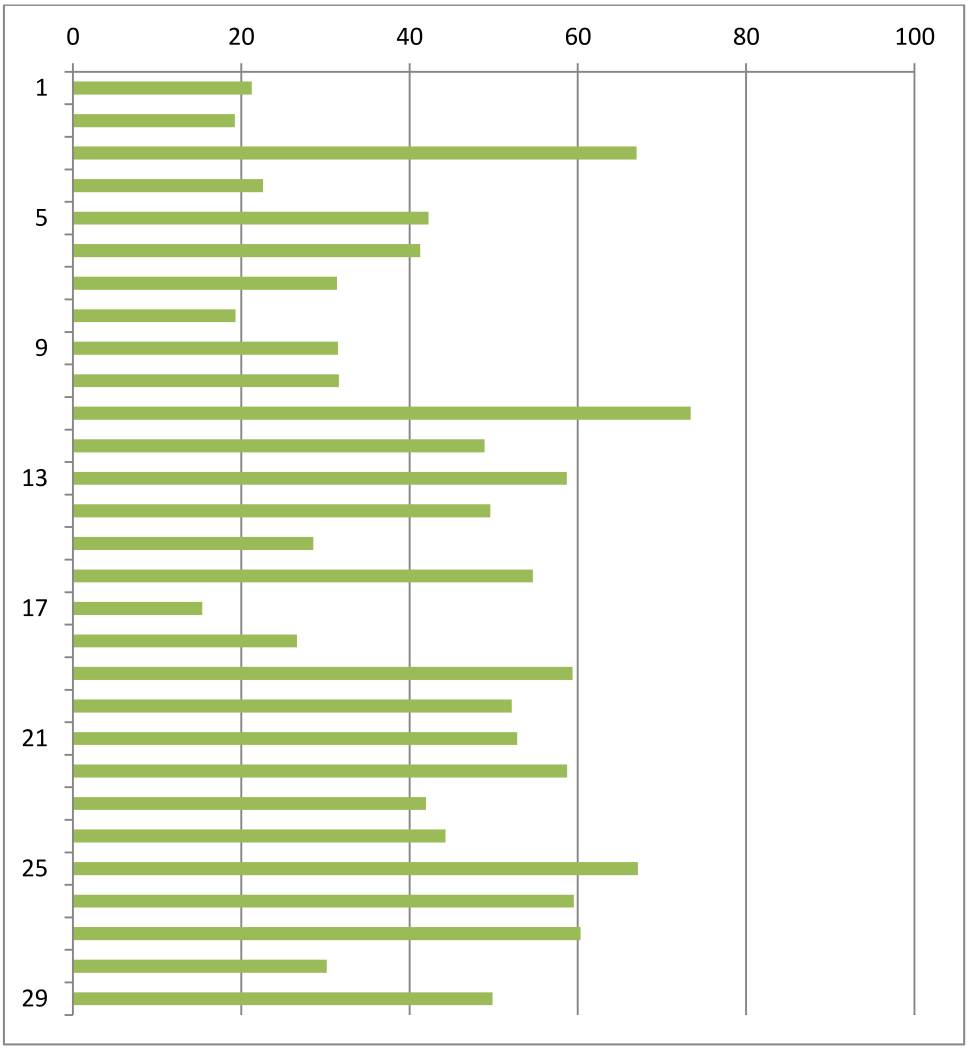
The following chart shows the results of the participants' judgments of the questions' relevance. According to this study, the questions 1 and 8 were rated clearly below average. Question 1 was "How to create accurate personalized assessment relevant to each student's progress of learning content?"; question 8 was "How to vary instructional strategies to increase learning efficiency?". A reason for this result might be the fact that questions related to a change of a teacher's own strategies and materials were less liked and thus less rated by the participating teachers. The highest rated question was 11, "How can teachers save time while creating individualized development/learning plans?", most likely the saving of teachers' own time was the reason for this. Similarly, questions referring to the efforts of personalizing teaching were ranked highest. This result provides some proof for the fundamental ideas of Learning Analytics in general and Lea's Box in particular.



The next chart shows the participants' judgements of the quality, appropriateness, and completeness of the solutions provided by the Lea's Box system (also rating on a scale from 1 to 10). The lowest rated solutions referred to the questions 2, 4, 15, and 17. Question 2 ("How to assess progress? How to assess value added by school/teacher?") referred to the means of assessment which hasn't been targeted by the project directly, also question 4 ("How does information disseminated by peers as well as peer-to-peer learning effect knowledge acquisition?"), question 15 ("What materials / topics should individual students attend – what homework would be the most effective for an individual?"), and question 17 ("Which social standing do students have in a group?") haven't been in the very focus of the project. Specifically peer ratings as well as social networks were out of the scope of the project. The highest ratings were given for questions 3, 12, and 25. We believe that the rating for question 3 ("How can students/teachers meaningfully reflect on student work to create higher quality work?") is based on the excellent appraisal of the flower app. For this tool we received extremely positive feedback and it will be a key part of the exploitation beyond the project (in form of a respective product). The high ranking for question 12 (“Who are the best/good and who are the low performers?”) was due to the simple fact that this question can be answer with a large number of simple charts (e.g., those of the OLM). Although there are more tailored tools for the risk prediction (such as the Learning Performance Vector, LPV), we believe that the high results for question 25 (“Is there a risk for particular students to fail in a course?”) are also due to the fact that teachers can make (or at least believe to can) such predictions on the basis of simple performance charts.



As shown in the following chart, a trade-off index between relevance and quality of the solution is likely more interesting and significant to evaluate the results of the project. Thus we calculated the product of relevance (in the paper/pencil version only, to involve the same participants) and assessment; this results i a value from 0 to 100 where a high value indicates a high importance and a high quality of the related solution. As shown below, we received very good results for questions 3, 11, and 25. Question 3 (“How can students/teachers meaningfully reflect on student work to create higher quality work?”) yielded a high relevance (7,76) and a high quality (8,63). We assume that this result is due to the flower app as well as the OLM as a method to compare different sources of evaluation. Question 11 (“How can teachers save time while creating individualized development/learning plans?”) has a very high relevance (9,30) and solutions were judged high as well (7,89). We assume that the nature of CbKST-type structures conveyed a stronger and clearer the impression of the nature / structure of a learning domain and the possible learning paths. Equally, the decomposition of learning steps into fine grained competencies appeared being a promising solution for teachers. Finally, question 25 (“Is there a risk for particular students to fail in a course?”) also yielded both, high relevance (8,07) and high quality of the solution (8,33). Again, we do believe that this result is due to the possibility to give an answer the is comparably simple in the combination with the novel CbKST approach.



From a critical perspective, we also found some strong discrepancies between relevance and solutions. Again, for question 4 (“How does information disseminated by peers as well as peer-to-peer learning effect knowledge acquisition?”) we found a relevance of 7,07 and a solution quality of only 3,20. As mentioned before, the reason is that the involvement of peer ratings was out of the scope of the project as it evolved of the M34 months. Also for question 6 ("Which assessments measure best?") we found such strong discrepancy (8,63 vs 4,78). The reason here is that answer this question actually would require and external criterion. A solution from the project would be a comparison of data sources and different activities in the OLM. This features, apparently, was acknowledged by the teachers well enough. One feedback from studies is clearly, that the calculation of the learner model in the OLM and the visualization (via drilling down in the tree map) are too complicated and features are not user friendly enough. Finally, question 15 ("What materials / topics should individual students attend – what homework would be the most effective for an individual?") yielded diverging results (8,26 vs 3,46). The question, obviously, is highly important, the current setup of the Lea's Box does not provide sufficient links to teacher and learning materials. This is a clear indication for future extensions.

Main Conclusions

- The flower app appeared to be a great and usable tool to support teachers. It offers a user friendly interface to compare self-assessments, teacher evaluations and test results. This allows not only facilitating self-reflection of students and teachers, it also provides a neat basis for negotiations and for communicating results.
- Although the concept is disputed in the research community, the influence of peers is considered of high importance. On the one hand, competition is a strong driver of learning motivation, on the other hand it can strengthen self-esteem or the negotiation position of a student's opposed to the teacher. In the interaction with teachers often the analogon to competition in sports arises, where in all our experience the positive aspects prevail. Besides that, peers are seen as a factor strongly influencing academic performance by social interactions.
- The structural approach provided by CbKST provides useful information and – in simplified versions – are liked and understood by teachers. Teachers very much appreciate the idea that analytics are not purely based on statistical information but have an inherent "top-down" component.
- A link to concrete materials for teaching and learning were rated highly important. A clear recommendation for future developments is to integrate such features in the system. This might be realized by portfolios of learning objects, exercises, test items, etc. It must be possible to associate these materials with the competencies and competency models. This must be possible in a user-friendly manner. A first simple step in this direction was the mind mapping scenario of year 1. This can serve as a role model.
- The catalogue of pedagogical questions was an excellent starting point for designing, developing, and reporting on educational technology. Having the teacher responses, this is a valuable asset for researchers and developers far beyond the project. Thus we provide this catalogue as one prominent outcome of the project.

4.3. Pilot "Grazer Schulschwestern" – Study 1

In the following, we describe the results of the use case "Schulschwestern", named after the according school in Graz where two studies have been carried out. The first study focused on the topic of "Applied Biology: Ecosystems" from February to April 2016 and the second study was on the topic "Perception".

It became 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 was not possible to have a control group, due to the fact that the teacher could not split his class into two parts (neither locally nor treatment-wise).

Participants

The teacher who participated 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. He also participated I prior projects as part of a test-bed and is therefore already familiar with the platform and its tools (see below).

The class started with 34 students whereas one male student left the school in the middle of the semester. Initially there were 24 female and 10 male students with an average age of M=15.06 (SD=0.78) years, ranging from 14 to 17 years. One male student dropped out the class in the middle of the semester.

Materials

The following questionnaires and inventories have been filled out by the teacher:

- Learning Resource Evaluation Form
- Assessment of the student´s performances via myClass

Ad 1.) Learning Resource Evaluation Form

Overall, 55 learning resources (such as text documents, URLs, videos, exercises or even experiments) on the domain "Applied biology: ecosystems" have been collected in advance by the teacher. These learning resources have been evaluated by the teacher with regards to the 4 bipolar scales of the Index of Learning Styles Questionnaire from Soloman and Felder (for details see below): i) Sensing vs. intuitive, ii) visual vs. verbal, iii) active vs. reflective, and iv) sequential vs. global. As indicated by the "Learning Resource Evaluation Form" (see Appendix, section 0), the teacher received an instruction on the scales and had to evaluate each learning resource by a 5-point Likert scale on each of the four dimensions.

Ad 2.) Assessment of the student´s exercises via myClass

As described in more detail in section 0, on the one hand, students engaged with "digital" learning resources via their tablets, on the other hand, they had to carry out exercises in a paper-and-pencil format. These exercises have been related to a set of 14 competences, in line with the competence models defined by the Austrian ministry for education (Bundesministerium für Bildung und Frauen, 2012). The teacher evaluated the student's exercises by an 11 point Likert scale, from 0 to 10. These ratings, together with the relationships to the competences served as input for evaluating the student's competences by using the myClass tool (by normalized values, i.e. between 0 and 1).

The following questionnaires and inventories have been filled out <u>by the students</u>:

1. Index of Learning Styles Questionnaire
2. Knowledge Tests on the domain "Applied Biology: Ecosystems".
3. Learning Log-book
4. Performance self-assessments via myClass

Ad 1.) Index of Learning Styles Questionnaire

The Learning Styles Questionnaire from Soloman and Felder is based on the Felder–Silverman Learning/Teaching Style Model (Felder & Silverman, 1988). The Felder–Silverman Learning Model is described in Deliverable D3.5. The questionnaire consists of the following 4 bipolar scales:

- **Sensing** (concrete thinker, practical, oriented towards facts and procedures) or **intuitive** (abstract thinker, innovative, oriented toward theories and underlying meanings);
- **Visual** (prefer visual representation of presented material, such as pictures, diagrams and flow charts) or **verbal** (prefer written and spoken explanations);
- **Active** (learn by trying things out, enjoy working in groups) or **reflective** (learn by thinking things through, prefer working alone or with a single familiar partner);
- **Sequential** (linear thinking process, learn in small incremental steps) or **global** (holistic thinking process, learn in large steps).

Each scale encompasses 11 items, consisting of statements with 2 alternative responses to choose from. The original version of the Index of Learning Styles Questionnaire from Soloman and Felder can be found at https://www.engr.ncsu.edu/learningstyles/ilsweb.html and the German version is provided in the appendix. One item, item number 13 from the original version (In classes I have taken…(a) I have usually gotten to know many of the students OR (b) I have rarely gotten to know many of the students) has been excluded for the German version since it doesn´t reflect Austrian´s school reality: There are fixed classes in Austria.

Ad 2.) Knowledge Tests on the domain "Applied Biology: Ecosystems"

To measure the student´s declarative knowledge states before and after using the LEAs Box and weSPOT tools, a knowledge test on the domain "Ecosystems" has been provided to the students. A parallel version *A* of the test has been provided to half of the students as pre-test and for the other half as a post-test and vice versa for the second parallel test version *B*. Both parallel tests consisted of 14 Multiple-choice items with 4 choices to choose from.

Ad 3.) Learning Log-book

By exploring the learning domain via the FCA tool the students had to opportunity engage in a self-directed manner with those learning resources which have been evaluated by the teacher with respect to the Felder–Silverman Model (see Learning Resource Evaluation Form above). Self-directed in this context means, that the students selected learning resources to engage with by their own, and that they decided by their own on how much time to spend on a particular learning resource. The students were asked to fill out the so called "Learning Log-book" (see Appendix): One row for one learning resource, and for each learning resource, the date of its selection, its title, a short summary of its content and finally an evaluation of the learning resource, indicated by values between 0 and 5. This evaluation was with regards to the "helpfulness" of this particular learning resource for the student at his or her current learning state. The instruction for the log-book pointed at the difference between this evaluation aspect and an evaluation of the learning resource per se (e.g. a nice looking video about abiotic environmental components might be a "good" learning resource, but not necessarily "helpful" for beginners).

Ad 4.) Students self-evaluations of the Exercises via myClass

Similar as for the teacher (see above), students used the myClass tool to evaluate their own exercises and competences. The self-assessments have been compared with the teacher´s assessments and these comparisons were the basis for discussions (primarily in cases of significant differences between the two assessments) between teachers and students.

TOOLSET

The set of tools encompassed:

1. weSPOT environment,
2. myClass environment,
3. FCA tool,
4. Flower app.

Ad 1.) The weSPOT environment

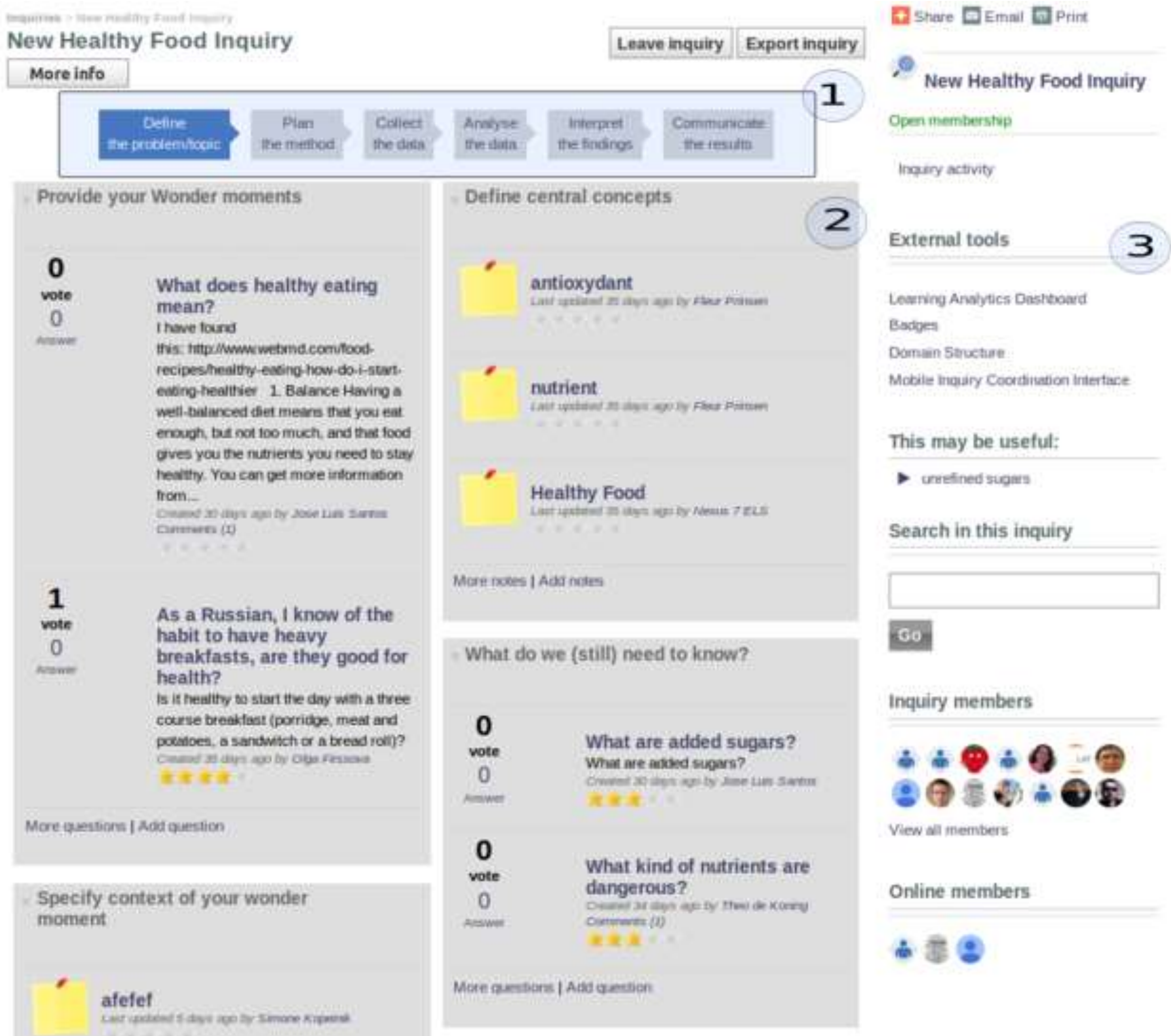
In the course of the weSPOT project (Working Environment with Social and Personal Open Tools), a platform which aims to support inquiry-based learning has been created (http://inquiry.wespot.net/). Inquiry-based learning (IBL) as pedagogical backbone regards inquiries as learning experiences, in which 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 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 the following table.

weSPOT´s toolkit

Tool	Description
weSpot Inquiry space engine (WIE)	WIE is a web-based platform 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 tool is shown in the following figure.



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 D2.6 (System Release III).

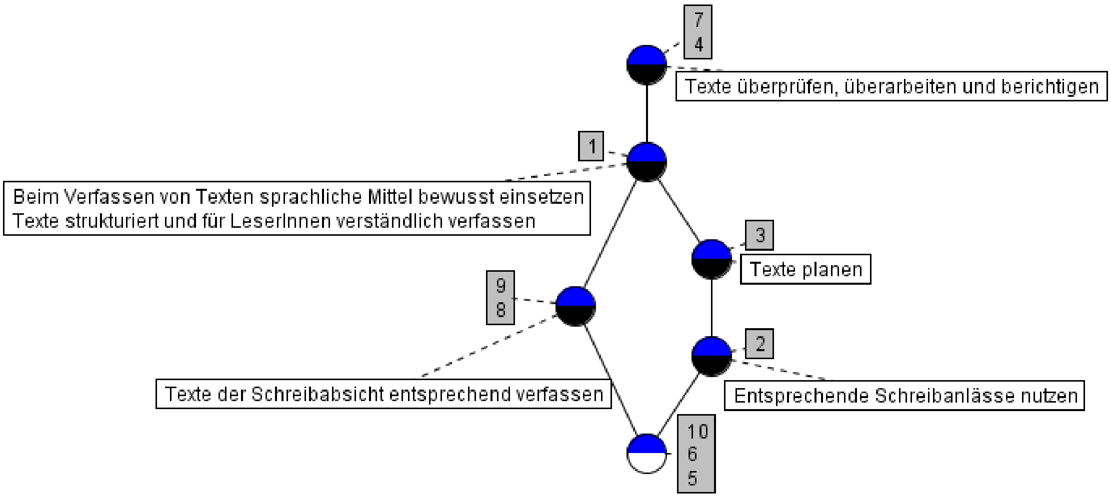
Ad 3.) The FCA tool

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, teachers 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.

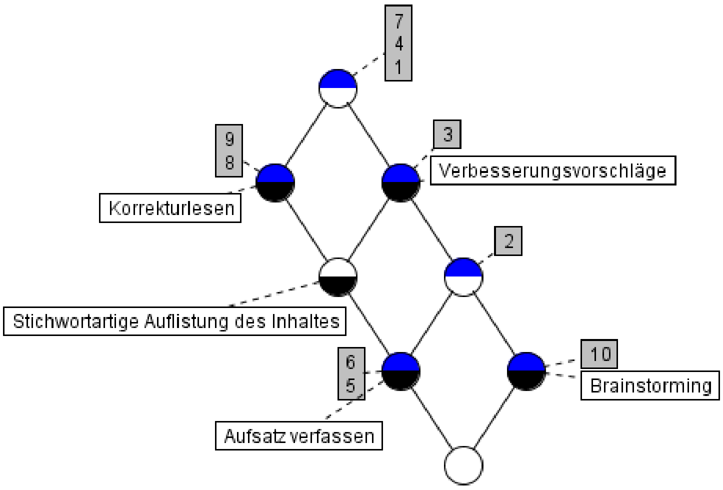
ii) A concept lattice with learners as attributes and competences as objects give an overview on the competence states of all students. An example is shown in the next figure.



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

In the figure above, 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 the next figure.



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 **Fehler! Verweisquelle konnte nicht gefunden werden.** can be similar interpreted as the one shown in the figure above: 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 App

The "Flower app" (for details see D3.4 Final release of LA/EDM services and algorithms) has been used to feed-back the test results, the teacher´s assessments, etc. to the students. In this sense, the Flower app as been applied for Open Learner Modelling.

PROCEDURE

The overall procedure can be divided in three consecutive phases:

Phase 1: Pre-tests and Preparation phase

The teacher sets up the inquiry in the platform (e.g. the inquiry phases, the objects, attributes and learning resources via the FCA tool). The topic of the inquiry is the domain “Applied biology: ecosystems”. The learning resources are evaluated with regards to the 4 bipolar scales of the Index of Learning Styles Questionnaire from Soloman and Felder.

Together with members of TUGraz, the teacher sets up the LEA´s BOX platform, i.e. the students, the competences, the activities and their relationships via the configuration tool.

The students fill out the first part of the declarative Knowledge Test (pre-test) and the Index of Learning Styles Questionnaire before they first enter the platform.

Phase 2: Intervention and Intermediate phase

The teacher uses the myClass tool 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 facilitates him to support his students individually, by getting an overview on their strengths and weaknesses.

The students interact with the environment and work on their inquiries, partly in groups and partly individually. For this, they use their tablet PCs. Their inquiry-related activities will be tracked by LARAe tool. The FCA tool gives an overview on the domain, i.e. to objects and attributes and their relationships, as well as the learning resources provided by the teacher. Students can decide by their own of which and how many learning resources they consume within the sessions. Once they are finished with one learning resource, they fill out the Learning Log-book. In addition to that, they work on exercises in paper-and-paper format.

Phase 3: Post-tests

The students fill out the second part of the declarative Knowledge Test (post-test) and deliver their Learning Log-books to the teacher. They evaluate their own performance via the myClass tool. The students´ self-evaluations and the teacher’s evaluations are visualized via the Flower app. These two assessments can be compared and discussed with the teacher (in the sense of a negotiated Open Learner Model).

RESULTS

In the following, we present the descriptive statistics, correlative and interference statistical results.

Descriptive Results

The following table summarizes the descriptive statistics, i.e. means, standard deviations, empirical range with minimal and maximal values, theoretical minimal and maximal values, and the number of observations (data points) for the following variables:

i) the four scales of the Index of Learning Styles Questionnaire which have been normalized to the range between 0.00 and 1.00,

ii) the absolute values for the knowledge pre- and post-tests,

iii) the normalized values knowledge pre- and post-test as well as the difference between those two values (called "Knowledge increase – normalized"),

iv) the number of exercises carried out by the students, the evaluation of these exercises by the teacher and the students´ self-evaluations of the exercises,

v) the number of consumed learning resources, the students' evaluation of these learning resources, and finally

vi) the number of activities tracked by the LARAe tool.

Descriptive statistics (with means, standard deviations, range) of the student's learning styles, declarative knowledge and learning activities

Variable	Mean (*Standard Deviation*)	Range: Min. empirical (*theoretical*)	Range: Max. empirical (*theoretical*)	*N*
Learning Styles: Active-Reflective	0,47 (*0,18*)	0,10 (*0,00*)	0,80 (*1,00*)	32
Learning Styles: Sensing-Intuitive	0,41 (*0,23*)	0,09 (*0,00*)	0,91 (*1,00*)	32
Learning Styles: Visual-Verbal	0,34 (*0,24*)	0,00 (*0,00*)	0,82 (*1,00*)	32
Learning Styles: Sequential-Global	0,39 (*0,18*)	0,09 (*0,00*)	0,80 (*1,00*)	32
Knowledge pre-test (absolute)	6,48 (*1,69*)	3,50 (*0,00*)	9,50 (14,00)	25
Knowledge post-test (absolute)	8,56 (*1,56*)	5,50 (*0,00*)	12,50 (14,00)	34
Knowledge pre-test (normalized)	0,46 (*0,12*)	0,25 (*0,00*)	0,68 (*1,00*)	25
Knowledge post-test (normalized)	0,61 (*0,11*)	0,39 (*0,00*)	0,89 (*1,00*)	34
Knowledge increase (normalized)	0,15 (*0,16*)	-0,14 (*0,00*)	0,43 (*1,00*)	25
Number of Exercises	7,85 (2,03)	0,00 (*0,00*)	10,00 (*10,00*)	34
Teacher´s evaluation of the exercises	0,722 (*1,00*)	0,00 (*0,00*)	1 (*1,00*)	34

Student´s self-evaluation of the exercises	0,544 (*1,00*)	0,00 (*0,00*)	1 (*1,00*)	24
Number of consumed Learning Resources	10,89 (*4,42*)	4 (*0,00*)	19 (55)	27
Learning Resource Evaluation	3,22 (*0,71*)	1,46 (*0,00*)	4,88 (*5,00*)	27
Number of weSPOT Activities	9,42 (*5,33*)	0 (*0,00*)	21 (-)	33

Correlative Results

The following table summarizes the bi-variate correlations (Pearson´s coefficients), and the amount of data-pairs (indicated by N) between the following variables:

i) the four scales of the Index of Learning Styles Questionnaire which have been normalized to the range between 0.00 and 1.00,

ii) the normalized values of the knowledge pre- and post-test as well as the difference between those two values (called "Knowledge increase"),

iii) the number of exercises carried out by the students, the evaluation of these exercises by the teacher and the students´ self-evaluations of the exercises,

iv) the number of consumed learning resources, the students' evaluation of these learning resources, and finally

v) the number of activities tracked by the LARAe tool.

Statistically significant correlations are highlighted by the grey colour of the according cells. Correlation coefficients marked by a "*" are statistically significant (p <.05), correlation coefficients marked by "**" are highly statistically significant (p < .001).

Correlation coefficients between the student's learning styles, declarative knowledge and learning activities

	Learning Styles: Active-Reflective	Learning Styles: Sensing-Intuitive	Learning Styles: Visual-Verbal	Learning Styles: Sequential-Global	Knowledge pre-test (normalized)	Knowledge post-test (normalized)	Knowledge increase (normalized)	Number of excercises	Teacher´s evaluation of the exercises	Student´s self-evaluation of the exercises	Number of consumed Learning Resources	Learning Resource Evaluation
Learning Styles: Sensing-Intuitive	0,009 N = 32											
Learning Styles: Visual-Verbal	0,141 N = 32	0,052 N = 32										

Learning Styles: Sequential-Global	-0,006 N = 32	0,323 N = 32	0,220 N = 32									
Knowledge pre-test (normalized)	-0,132 N = 25	-0,067 N = 25	0,024 N = 25	-0,082 N = 25								
Knowledge post-test (normalized)	-0,010 N = 32	0,058 N = 32	0,147 N = 32	-0,290 N = 32	0,067 N = 25							
Knowledge increase (normalized)	0,123 N = 25	-0,050 N = 25	0,129 N = 25	-0,039 N = 25	-0,733** N = 25	0,629** N = 25						
Number of exercises	0,196 N = 32	-0,053 N = 32	0,341 N = 32	0,078 N = 32	0,261 N = 25	0,297 N = 33	0,127 N = 25					
Teacher´s evaluation of the exercises	0,144 N = 32	-0,256 N = 32	-0,314 N = 32	-0,224 N = 32	-0,017 N = 25	-0,233 N = 34	0,128 N = 25	0,246 N = 33				
Student´s self-evaluation of the exercises	0,218 N = 32	0,063 N = 32	-0,234 N = 32	0,001 N = 32	-0,208 N = 25	-0,362* N = 34	-0,124 N = 25	-0,482** N = 33	-0,041 N = 155			
Number of consumed Learning Resources	0,250 N = 27	-0,014 N = 27	0,157 N = 27	-0,122 N = 27	0,228 N = 21	0,380* N = 27	0,131 N = 21	0,327 N = 27	0,537** N = 27	-0,377 N = 27		
Learning Resource Evaluation	0,284 N = 27	0,005 N = 27	-0,008 N = 27	-0,160 N = 27	0,233 N = 21	-0,071 N = 27	-0,338 N = 21	0,058 N = 27	0,409* N = 27	-0,282 N = 27	0,255 N = 27	
Number of weSPOT Activities	0,151 N = 32	-0,069 N = 32	0,206 N = 32	0,025 N = 32	0,096 N = 25	0,093 N = 33	0,083 N = 25	0,658** N = 33	0,427* N = 33	0,672** N = 33	0,512** N = 27	0,177 N = 27

As indicated in the table, the students' knowledge increase correlates with the knowledge per-test and post-tests results. This is not surprising, since the knowledge increase values are derived from the other two values. Thus, these correlation coefficients shouldn´t be interpreted too much. However, the negative coefficient between knowledge increase and the knowledge pre-test hints at the rather trivial result, that the lower the pre-test values, the higher the “potential for improvement”.

The higher (or the better) the student´s self-evaluations of their exercises, the lower the students’ knowledge post-test results and the lower the number of exercises carried out by the students. A possible explanation concerning the latter relationship could be that those students who submitted fewer exercises had the subjective impression that they delivered higher quality, rather than just quantity.

The higher the number of consumed learning resources, the higher the knowledge post-test results and the teachers average evaluation of the student´s exercises. Put simple: the students learned something by consuming and interacting with the learning resources. This seems to be a trivial result; however, we consider it as an important one. The number of activities tracked by the LARAe tool reflects the i) number of exercises submitted to the teacher, ii) the teacher´s evaluation of these exercises, iii) the students´ self-evaluation of these exercises and iv) the number of consumed learning resources. All of these relationships are reflected by positive correlation coefficients. The first relation shouldn´t be too much interpreted, since the number of exercises submitted to the teacher contributes to the number of activities. However, all other relations are not that direct and thus, more interesting. It means that the LARAe tool enables to predict other activity-related achievements of the students.

The following analysis aims to contribute with a small jigsaw puzzle-piece to the famous matching hypothesis in the field of learning styles (for details see Deliverable 3.5). In a nutshell, the matching-hypothesis suggests that the instructional style or the nature of the learning resources to be used (e.g. “concrete” or “visual”) should be aligned with the learner´s dominant learning style. In order to contribute at least a little bit to this still unresolved research question, we tried to “measure” the matching between the students learning styles and the nature of the learning resources by the following approach: As mentioned above, all learning resources have been evaluated by the teacher with regards to the 4 bipolar scales of the Index of Learning Styles Questionnaire. The normalized values of these evaluations deliver a “vector” with 4 numbers between 0 and 1. A similar vector can be defined by the normalized learning style values of the students. For each factor (or dimension) we computed the absolute difference between the “learner´s learning style value” and the “learning resource´s learning style value”. The lower this absolute difference, the higher the “matching” between a particular student and the learning resource. These absolute differences values have been correlated with the students evaluation of the learning resources (delivered via the Learning Log-Book). The results of this analysis are shown in the following table.

As indicated by the correlation coefficients, there is no relation between our operationalization of „matching“ between students and learning resources and the students evaluation of these learning resources. Potential explanations for these rather surprising results will be outlined in the discussion section.

Correlation coefficients between the "learner-learning resource matching" for the scales and overall values of the Index of Learning Styles and learning resource evaluation (study 1)

	Active –Reflective Dimension: Learner – Learning Resource Difference	Sensing-Intuitive Dimension: Learner – Learning Resource Difference	Visual-Verbal Dimension: Learner – Learning Resource Difference	Sequential-Global Dimension: Learner – Learning Resource Difference	Index of Learning Styles: Learner – Learning Resource Difference
Learning Resource Evaluation	0,016 N = 206	0,021 N = 206	0,002 N = 206	-0,037 N = 206	0,004 N = 206

Interference Statistics

In the following, we will apply some Analysis of Variance to check if the "independent variables" i) gender and/or ii) the pre-test knowledge have some (causal) effect on the following 8 "dependent variables":

i) the normalized knowledge post-test values,

ii) the relative increase of the knowledge ("Knowledge increase – normalized"),

iii) the number of exercises carried out by the students,

iv) the evaluation of these exercises by the teacher,

v) the students´ self-evaluations of the exercises,

vi) the number of consumed learning resources,

vii) learning resource evaluation, and

viii) the number of activities tracked by the LARAe tool.

For the second factor, i.e. the pre-test knowledge, we divided the students into two groups, into the group "low pre-test knowledge" and the group "high pre-test knowledge" based on the median value.

From a methodological point of view, a 2x2 Multivariate Analysis of Variance with 7 dependent variables would be the appropriate statistical method. However, there are several data-related issues which are crucial for such an analysis: there is a non-equal distribution with regards to both independent variables. In the overall sample, there are 24 female and 10 male; in the sub-sample for which knowledge pre-test values are available, there are 18 female and 7 male. The median-split didn´t result in perfectly equal sub-sample sizes: Based on the median-split, 10 students are part of the "low pre-test knowledge" group, and 15 students ate part of the "high pre-test knowledge" group. Since these two independent variables are not completely uncorrelated with each other, the distribution of the students with the 2x2 cells is even less equal: 6 female students would be part of the "low pre-test knowledge" group and 12 female students would be part of the "high pre-test knowledge" group; whereas 4 male students would be part of the "low pre-test knowledge" group and 3 male students would be part of the "high pre-test knowledge" group. When considering only the complete data-sets, it is getting even worse (e.g. only one male students remains). This makes the analysis of interactions between gender and pre-knowledge level generally impossible. This argument, together with the lack of complete datasets of male students with respect to the 7 dependent variables) is the reason, why we decided to apply a series of analyses of variance (ANOVA´s), separated for "gender" and for pre-knowledge level (even without Bonferroni-correction). Thus, the results of this "pragmatic approach" should be interpreted with great caution or not at all, respectively (again from a pragmatic point of view) only as hints for potential research questions and further studies.

Results of the univariate comparisons between female and male students

Source: Gender	Distribution			
Dependent Variable	female	male	*F*	*p*
Knowledge post-test (normalized)	24	10	0,385	0,539
Knowledge increase (normalized)	18	7	0,657	0,426
Number of Exercises	23	10	3,038	0,091
Teacher´s evaluation of the exercises	23	10	6,080	0,019*
Student´s self-evaluation of the exercises	22	2	0,441	0,513
Number of consumed Learning Resources	20	7	5,759	0,024*
Learning Resource Evaluation	20	7	1,076	0,309
Number of weSPOT Activities	23	10	15,819	0,000**

Female students received higher or better evaluations of their exercises by the teacher (*M* = 0,74; *SD* = 0,07) than male students (*M* = 0,67; *SD* = 0,06). Female students consumed more learning resources (*M* = 12,00; *SD* = 4,38) than male students (*M* = 7,71; *SD* = 2,87). Finally, the same can be said about the number of weSPOT activities: female students carried out more of these activities (*M* = 11,43; *SD* = 4,56) than male students (*M* = 4,89; *SD* = 3,74).

As mentioned in the introduction of this section, due to methodological reasons, the results should be interpreted with great caution (if interpreted at all). However, when discussing these results with the teacher he was not surprised at all about these results and mentioned that he expected results like these (at least in this particular classroom, not generally speaking).

Results of the univariate comparisons between students with low and high pre-test knowledge

Source: knowledge pre-test	Distribution			
Dependent Variable	low pre-knowledge	high pre-knowledge	*F*	*p*
Knowledge post-test (normalized)	10	15	0,003	0,958
Knowledge increase (normalized)	10	15	14,183	0,001**
Number of Exercises	10	15	1,197	0,285
Teacher´s evaluation of the exercises	10	15	0,089	0,768
Student´s self-evaluation of the exercises	6	12	0,113	0,741
Number of consumed Learning Resources	7	14	0,927	0,348
Learning Resource Evaluation	7	14	0,615	0,443
Number of weSPOT Activities	10	15	0,219	0,645

As already indicated by the correlative analysis, the lower the pre-test values, the higher the “potential for improvement”. Students with lower knowledge pre-test levels have a higher knowledge increase at the end the study (*M* = 0,27; *SD* = 0,08) than students with higher knowledge pre-test levels (*M* = 0,08; *SD* = 0,14).

Discussion

With regards to the descriptive statistics, we would like to point out at the medium means and the rather high range (indicated by the minimum and maximum values as well as the standard deviation) of all four scales of the Index of Learning Styles Questionnaire. A medium “difficulty” (i.e. a medium mean) is a main contributing factor for a questionnaires´ psychometric quality. It might be also an indication for the quality of the German version which has been translated because of to this study and its research questions

The most interesting correlative results are negative relationship between the student´s self-evaluations of their exercises and the students' knowledge post-test results. It seems that poorer students (whereas "poorer" in this context is solely indicated by the lower result of a declarative knowledge test at the end of the study) overestimate their achievements. Poorer students also carried out a lower number of exercises. A possible explanation could be that those students who submitted fewer exercises had the subjective impression that they delivered higher quality, rather than quantity.

Interestingly, the number of activities tracked by the LARAe tool reflects a wide range of other variables, such as the i) number of exercises submitted to the teacher, ii) the teacher´s evaluation of these exercises, iii) the students´ self-evaluation of these exercises and iv) the number of consumed learning resources.

The results on the "matching hypothesis" were surprising, to say the least. To be honest, we expected a positive correlation between the "learner-learning resource matching" and the evaluation of the learning resources of the students (in our context, we would have expected negative correlation coefficients: the lower the "distance", the better the learning resource evaluation values. Together with the teacher, we identified 4 potential explanations:

i) Plain simple, the results are as they are. There is just no relation between "learner-learning resource matching" and the evaluation of the learning resources. In other words: the matching hypothesis can be rejected.

ii) We asked the students to evaluate how "helpful" the learning resource was for them - at this particular point in time respectively at their current knowledge and learning state, rather than how much they liked or enjoyed this particular learning resource in general. Maybe this difference was a bit hard to understand, and it could be the case that the according instruction was confusing for some students. In consequence, this might have resulted in additional "noise" in the data.

iii) Maybe it's exactly this kind of evaluation aspect ("helpfulness") for which the matching hypothesis doesn´t hold. Since we haven´t asked for both mentioned aspects (i.e. also for how much they liked the learning resource) we cannot answer this question. However, it is from our point of view an interesting research question for the future.

iv) There were some learning resources which have been generally evaluated with either extremely high or low values, from all students, i.e. independent from their gender, pre-test knowledge or learning styles. Such "extreme example" together with low variance reduces the correlation coefficient.

The results of the Analysis of Variance series hint at significant gender differences. As mentioned above, the teacher was not surprised by these results. However, as also mentioned above, the results shouldn´t be over-interpreted due to statistical reasons.

4.4. Pilot "Grazer Schulschwestern" – Study 2

Participants, Materials

The teacher and the students were the same as in Study 1. The toolset which has been used by the teacher and the students was the same as for study 1.

Questionnaires and inventories

The following questionnaires and inventories have been filled out by the teacher:

1. Learning Resource Evaluation Form
2. Assessment of the student´s performances via myClass

Ad 1.) Learning Resource Evaluation Form

Overall, 19 learning resources (such as text documents, URLs, videos, exercises) on the domain "Perception" have been collected by the teacher. As for study 1, these learning resources have been evaluated by the teacher with regards to the 4 bipolar scales of the Index of Learning Styles Questionnaire from Soloman and Felder.

Ad 2.) Assessment of the student´s exercises via myClass

As for study 1, besides the "virtual activities", students also had to work on paper-and-pencil exercises. These exercises have been mapped on a set of 9 competences defined by the Austrian ministry for education (Bundesministerium für Bildung und Frauen, 2012). The teacher evaluated the student's exercises by an 11 point Likert scale, from 0 to 10). These ratings, together with the relationships to the competences served as input for evaluating the student's competences by using the myClass tool (by normalized values, i.e. between 0 and 1).

The following questionnaires and inventories have been filled out by the students:

1. Index of Learning Styles Questionnaire
2. Knowledge Tests on the domain "Perception".
3. Learning Log-book
4. Intrinsic Motivation Scale

This list of questionnaires and inventories is nearly identical with the one for study. However, in this study 2 we also applied the Intrinsic Motivation Scale to measure the students' motivational state at different points in time (see also procedure, section 0). Unfortunately, due to time constraints at the very end of the summer term 2016, it was necessary to skip the “Performance self-assessments via myClass”.

Ad 1.) Index of Learning Styles Questionnaire

The Index of Learning Styles Questionnaire has been already described above. The original version can be found at https://www.engr.ncsu.edu/learningstyles/ilsweb.html and the German version is provided in the appendix.

Ad 2.) Knowledge Tests on the domain “Perception”.

To measure the student´s declarative knowledge states before and after using the LEAs Box tools, a knowledge test on the domain “Perception” has been provided to the students. The pre-test consisted of 9 multiple choice items and the post-test consisted of 10 items.

Ad 3.) Learning Log-book

The Learning Log-book has been already described in section **Fehler! Verweisquelle konnte nicht gefunden werden.** and is provided in the appendix

Ad 4.) Intrinsic Motivation Scale

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 original version of IMI (in English) as well as the German version is provided in the appendix.

PROCEDURE

As for study 1, we divide the study into three consecutive phases:

Phase 1: Pre-tests and Preparation phase

The teacher sets up the inquiry in the weSPOT platform (e.g. the inquiry phases, the objects, attributes and learning resources via the FCA tool). The topic of the inquiry is the domain “Perception”. The learning resources are evaluated with regards to the 4 bipolar scales of the Index of Learning Styles Questionnaire from Soloman and Felder.

Together with members of TUGraz, the teacher sets up the LEA´s BOX platform, i.e. the students, the competences, the activities and their relationships in the configuration tool.

The students fill out the first part of the declarative Knowledge Test (pre-test) and the Intrinsic Motivation Inventory (IMI) before they enter the LEAs Box platforms.

Phase 2: Intervention and Intermediate phase

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.

The students interact with the environment and work on their inquiries, partly in groups and partly individually. For this, they use their tablet PCs. Their inquiry-related activities will be tracked by the LARAe tool. The FCA tool gives an overview on the domain, i.e. to objects and attributes and their relationships, as well as the learning resources provided by the teacher. Students can decide by their own of which and how many learning resources they consume within the sessions. Once they are finished with one learning resource, they fill out the Learning Log-book. Besides that, they also work on exercises in paper-and-paper format. At a medium stage of this phase, the students fill out IMI for the second time, as an intermediate test.

Phase 3: Post-tests

The students fill out the second part of the declarative Knowledge Test (post-test) and the IMI for the third time. They deliver their Learning Log-books to the teacher. Finally, they fill out the Index of Learning Styles Questionnaire for the second time at the very end of the summer term.

The teacher evaluated the student´s exercises and conducted a semi-structured Interview with the members of TUGraz.

RESULTS

Descriptive Results

The following table summarizes the descriptive statistics, i.e. means, standard deviations, empirical range with minimal and maximal values, theoretical minimal and maximal values, and the number of observations (data points) for the following variables:

i) the four scales of the Index of Learning Styles Questionnaire which have been normalized to the range between 0.00 and 1.00,

ii) the three scales of the Intrinsic Motivation Scale and the overall intrinsic motivation for the pre- intermediate- and post-test

iii) the absolute values for the knowledge pre- and post-tests,

iv) the normalized values knowledge pre- and post-test as well as the difference between those two values (called “Knowledge increase – normalized”),

v) the number of exercises carried out by the students and the evaluation of these exercises by the teacher,

vi) the number of consumed learning resources, the student’s evaluation of these learning resources, and finally,

vii) the number of activities tracked by the LARAe tool.

Descriptive statistics (with means, standard deviations, range) of the student’s learning styles, intrinsic motivation, declarative knowledge and learning activities

Variable	Mean (*Standard Deviation*)	Range: Min. empirical (*theoretical*)	Range: Max. empirical (*theoretical*)	*N*
Learning Styles: Active-Reflective	0,50 (*0,16*)	0,20 (*0,00*)	0,75 (*1,00*)	28
Learning Styles: Sensing-Intuitive	0,45 (*0,23*)	0,09 (*0,00*)	1,00 (*1,00*)	28
Learning Styles: Visual-Verbal	0,38 (*0,23*)	0,00 (*0,00*)	0,82 (*1,00*)	28
Learning Styles: Sequential-Global	0,38 (*0,17*)	0,09 (*0,00*)	0,73 (*1,00*)	28
Intrinsic Motivation pre-test	5,36 (*1,05*)	3,11 (*1,00*)	6,89 (*7,00*)	33
Interest pre-test	5,28 (*1,16*)	3,00 (*1,00*)	7,00 (*7,00*)	33
Effort pre-test	5,64 (*1,07*)	3,33 (*1,00*)	7,00 (*7,00*)	33
Value pre-test	5,15 (*1,32*)	2,33 (*1,00*)	7,00 (*7,00*)	33
Intrinsic Motivation intermediate-test	5,48 (*0,86*)	3,67 (*1,00*)	6,89 (*7,00*)	31
Interest intermediate-test	5,82 (*0,76*)	4,00 (*1,00*)	7,00 (*7,00*)	31
Effort intermediate-test	5,10 (*1,29*)	2,00 (*1,00*)	7,00 (*7,00*)	31
Value intermediate-test	5,52 (*0,86*)	3,67 (*1,00*)	7,00 (*7,00*)	31
Intrinsic Motivation post-test	5,02 (*1,05*)	3,22 (*1,00*)	6,67 (*7,00*)	29
Interest post-test	5,33 (*1,08*)	2,67 (*1,00*)	7,00 (*7,00*)	29
Effort post-test	4,72 (*1,50*)	2,00 (*1,00*)	7,00 (*7,00*)	29

Value post-test	4,99 (*1,11*)	2,33 (*1,00*)	6,67 (*7,00*)	29
Knowledge pre-test (absolute)	4,04 (*1,30*)	2,00 (*0,00*)	7,00 (*9,00*)	24
Knowledge post-test (absolute)	7,63 (*1,56*)	4,00 (*0,00*)	10,00 (*10,00*)	24
Knowledge pre-test (normalized)	0,45 (*0,14*)	0,22 (*0,00*)	0,78 (*1,00*)	24
Knowledge post-test (normalized)	0,76 (*0,16*)	0,40 (*0,00*)	1,00 (*1,00*)	24
Knowledge increase (normalized)	0,32 (*0,18*)	-0,04 (*0,00*)	0,58 (*1,00*)	16
Number of Exercises	7,15 (*1,44*)	4,00 (*0,00*)	9,00 (*9,00*)	33
Teacher´s evaluation of the Exercises	6,57 (*1,65*)	3,38 (*0,00*)	8,86 (*10,00*)	33
Number of consumed Learning Resources	9,13 (*4,01*)	3,00 (*0,00*)	16,00 (*19,00*)	23
Learning Resource Evaluation	3,46 (*0,66*)	2,00 (*0,00*)	5,00 (*5,00*)	23
Number of weSPOT Activities	8,44 (*5,18*)	0,00 (*0,00*)	20,00 (n.a.)	32

Correlative Results

The following table shows the correlation coefficients between the subscales of the Intrinsic Motivation Inventory and its overall value for the pre- intermediate and post-test.

Statistically significant correlations are highlighted by the grey colour of the according cells. Correlation coefficients marked by "*" are statistically significant (p <.05), correlation coefficients marked by "**" are highly statistically highly significant (p < .001).

Correlative results (test-retest reliability coefficients) for the scales of the Intrinsic Motivation Scale

	Interest (pre-test)	Effort (pre-test)	Value (pre-test)	Overall Intrinsic Motivation (pre-test)	Interest (intermediate-test)	Effort (intermediate-test)	Value (intermediate-test)	Overall Intrinsic Motivation (intermediate-test)	Interest (post-test)	Effort (post-test)	Value (post-test)
Effort (pre-test)	0,694** N = 33										
Value (pre-test)	0,661** N = 33	0,661** N = 33									
Overall Intrinsic Motivation (pre-test)	0,884** N = 33	0,876** N = 33	0,890** N = 33								
Interest	0,599**	0,669**	0,667**	0,714*							

(intermediate-test)	N = 31	N = 31	N = 31	N = 31							
Effort (intermediate-test)	0,598** N = 31	0,789** N = 31	0,674** N = 31	0,757** N = 31	0,784** N = 31						
Value (intermediate-test)	0,456** N = 31	0,531** N = 31	0,639** N = 31	0,606** N = 31	0,555* N = 31	0,590** N = 31					
Overall Intrinsic Motivation (intermediate-test)	0,632** N = 31	0,775** N = 31	0,752** N = 31	0,797** N = 31	0,878** N = 31	0,935** N = 31	0,798** N = 31				
Interest (post-test)	0,543* N = 29	0,446* N = 29	0,566* N = 29	0,597* N = 29	0,527* N = 27	0,531* N = 27	0,557* N = 27	0,626** N = 27			
Effort (post-test)	0,565* N = 29	0,587* N = 29	0,751** N = 29	0,731** N = 29	0,710** N = 27	0,775** N = 27	0,438* N = 27	0,773** N = 27	0,630** N = 29		
Value (post-test)	0,618** N = 29	0,497* N = 29	0,851** N = 29	0,761** N = 29	0,306 N = 27	0,447* N = 27	0,612* N = 27	0,532* N = 27	0,536* N = 29	0,577* N = 29	
Overall Intrinsic Motivation (post-test)	0,674** N = 29	0,609** N = 29	0,853** N = 29	0,822** N = 29	0,623* N = 27	0,708** N = 27	0,619* N = 27	0,770** N = 27	0,833** N = 29	0,897** N = 29	0,812** N = 29

As indicated in the table, the correlation coefficients between subscales and overall values of the Intrinsic Motivation inventory across the pre-, intermediate-, and post-tests are significant (and to a vast majority highly significant). Thus, for the remaining analysis, we will focus only on the Intrinsic Motivation overall values.

The following table summarizes the bi-variate correlations (Pearson´s coefficients), and the amount of data-pairs (indicated by N) between the following variables:

i) the four scales of the Index of Learning Styles Questionnaire which have been normalized to the range between 0.00 and 1.00,

ii) the normalized values of the knowledge pre- and post-test as well as the difference between those two values (called “Knowledge increase”),

iii) the number of exercises carried out by the students and the evaluation of these exercises by the teacher,

iv) the number of consumed learning resources, the student’s evaluation of these learning resources

v) the number of activities tracked by the LARAe tool, and finally,

vi) the Intrinsic Motivation overall values for the pre-, intermediate-, and post-test.

Correlation coefficients between the student’s learning styles, intrinsic motivation, declarative knowledge and learning activities

	Learning Styles: Active-Reflective	Learning Styles: Sensing-Intuitive	Learning Styles: Visual-Verbal	Learning Styles: Sequential-Global	Knowledge pre-test (normalized)	Knowledge post-test (normalized)	Knowledge increase (normalized)	Number of exercises	Teacher´s evaluation of the exercises	Number of weSPOT Activities
Learning Styles: Sensing-Intuitive	0,063 N = 28									
Learning Styles: Visual-Verbal	0,604* N = 28	0,151 N = 28								
Learning Styles: Sequential-Global	0,144 N = 28	0,487* N = 28	-0,030 N = 28							
Knowledge pre-test (normalized)	0,231 N = 19	0,480* N = 19	-0,119 N = 19	0,042 N = 19						
Knowledge post-test (normalized)	0,420 N = 22	-0,074 N = 22	0,088 N = 22	-0,246 N = 22	0,188 N = 16					
Knowledge increase (normalized)	0,047 N = 15	-0,154 N = 15	0,291 N = 15	0,162 N = 15	-0,670* N = 16	0,604* N = 16				
Number of exercises	-0,029 N = 28	0,076 N = 28	0,151 N = 28	-0,012 N = 28	-0,198 N = 23	0,276 N = 23	0,529* N = 15			
Teacher´s evaluation of the exercises	0,093 N = 28	-0,326 N = 28	0,026 N = 28	0,046 N = 28	-0,037 N = 23	0,076 N = 23	0,122 N = 15	0,337 N = 33		
Number of weSPOT Activities	0,193	-	0,123	0,11	-0,178	0,082	0,444	0,515*	0,52	

	N = 27	0,131 N = 27	N = 27	6 N = 27	N = 23	N = 22	N = 15	N = 32	1* N = 32	
Overall Intrinsic Motivation (pre-test)	-0,013 N = 28	-0,336 N = 28	-0,390* N = 28	-0,162 N = 28	0,094 N = 24	-0,008 N = 23	-0,183 N = 16	-0,001 N = 32	0,359* N = 32	0,182 N = 31
Overall Intrinsic Motivation (intermediate-test)	-0,348 N = 26	-0,504* N = 26	-0,525* N = 26	-0,252 N = 26	-0,174 N = 23	-0,188 N = 22	-0,171 N = 15	-0,008 N = 30	0,350 N = 30	0,157 N = 29
Overall Intrinsic Motivation (post-test)	-0,203 N = 28	-0,379* N = 28	-0,475* N = 28	-0,039 N = 28	-0,046 N = 20	-0,020 N = 23	-0,017 N = 16	0,135 N = 28	0,418* N = 28	0,365 N = 27

As indicated by the grey cells in the upper left corner of the correlation matrix, in this study we identified some statistically significant relationships between different learning styles dimensions: the more “verbal” the student´s learning style, the more “reflective” his or her style of learning. And vice versa, the more “visual”, the more “active” is the learning style. In addition to that, “intuitive” thinkers are more likely also more “global” thinkers, and vice versa, “sensing” learners are more likely “sequential” learners. Interestingly, this “sensing-intuitive” dimensions correlates also with some other achievement-related and motivational variables: “Intuitive” learners had higher knowledge pre-test results, but lower intrinsic motivation values during and at the end of the study. A similar relationship has been observed between “verbal” learners and intrinsic motivation: “verbal” learners are less motivated during the whole inquiry project, as indicated by lower pre- intermediate- and post-test values at the Intrinsic Motivation Inventory.

The remaining relationships are similar as for study 1: Not surprisingly, the lower the knowledge pre-test result, the higher the knowledge increase. And the higher the knowledge increase the higher the knowledge post-test results.

As for study 1, the number of exercises carried out by the students the higher the knowledge increase.

The higher the students intrinsic motivation (at least at the pre-test and the post-test point in time) and the more activities observed ad tracked vie the LARAe tool, the better the teacher´s evaluation of the students´ exercises.

The following table shows the results on the “matching hypothesis” between the students´ learning styles and the nature of the learning resources.

Correlation coefficients between the "learner-learning resource matching" for the scales and overall values of the Index of Learning Styles and learning resource evaluation (study 2)

	Active –Reflective Dimension: Learner – Learning Resource Difference	Sensing-Intuitive Dimension: Learner – Learning Resource Difference	Visual-Verbal Dimension: Learner – Learning Resource Difference	Sequential-Global Dimension: Learner – Learning Resource Difference	Index of Learning Styles: Learner – Learning Resource Difference
Learning Resource Evaluation	0,520 N = 146	0,000 N = 158	0,000 N = 158	0,000 N = 158	0,000 N = 158

As for study 1, there are no relations between our operationalization of the „matching" between students and learning resources and the students evaluation of these learning resources. Potential explanations for these rather surprising results have been already outlined.

Interference Statistics

In the following, we report the results of a series of Analysis of Variance to check if the "independent variables" i) gender, ii) the pre-test knowledge and iii) the level of intrinsic motivation at the pre-test have some (causal) effect on the following 7 "dependent variables":

i) the normalized knowledge post-test values,

ii) the relative increase of the knowledge ("Knowledge increase – normalized"),

iii) the number of exercises carried out by the students,

iv) the evaluation of these exercises by the teacher,

v) the number of consumed learning resources,

vi) learning resource evaluation, and

vii) the number of activities tracked by the LARAe tool.

For both, the second and third factor, i.e. the pre-test knowledge and pre-test intrinsic motivation, we divided the students into two groups based on a median spilt.

As outlined in the introduction above, due to statistical problems of the datasets (e.g. small numbers of datasets, unequal distribution of students between the groups, etc.) the results should be either interpreted with great caution or not at all. We apply a pragmatic approach, and consider the results as potential hints for research questions to be considered in further studies.

Results of the univariate comparisons between female and male students

Source: Gender	Distribution			
Dependent Variable	female	male	*F*	*p*
Knowledge post-test (normalized)	18	6	0,049	0,826
Knowledge increase (normalized)	11	5	3,275	0,092
Number of Exercises	24	9	16,880	0,000**
Teacher´s evaluation of the exercises	24	9	9,953	0,004*
Number of consumed Learning Resources	19	4	3,883	0,062
Learning Resource Evaluation	19	4	5,584	0,028*
Number of weSPOT Activities	23	9	15,635	0,000**

The gender effects summarized in **Fehler! Verweisquelle konnte nicht gefunden werden.** table are very similar as the one for study 1. Female students received higher or better evaluations of their exercises by the teacher (*M* = 7,05; *SD* = 1,63) than male students (*M* = 5,26; *SD* = 0,76). Female students carried out more activities (*M* = 10,30; *SD* = 4,28)

than male students (*M* = 3,67; *SD* = 4,24). In addition to that, female students did more exercises (*M* = 7,67; *SD* = 1,17) than male students (*M* = 5,78; *SD* = 1,20).

Interestingly, male students evaluated the learning resources more favorable exercises (*M* = 4,12; *SD* = 0,73) than female students (*M* = 3,32; *SD* = 0,58).

As indicated by the following table, there are no statistically significant differences between students with low pre-test knowledge and students with high pre-test knowledge with respect to the 7 dependent variables.

Results of the univariate comparisons between students with low and high pre-test knowledge

Source: Knowledge pre-test	Distribution			
Dependent Variable	low pre-knowledge	high pre-knowledge	*F*	*p*
Knowledge post-test (normalized)	11	5	0,795	0,388
Knowledge increase (normalized)	11	5	4,140	0,061
Number of Exercises	16	7	2,097	0,162
Teacher´s evaluation of the exercises	16	7	0,020	0,890
Number of consumed Learning Resources	11	6	0,825	0,378
Learning Resource Evaluation	11	6	0,020	0,889
Number of weSPOT Activities	16	7	0,000	0,994

Results of the univariate comparisons between students with low and high intrinsic motivation at the pre-test

Source: Intrinsic Motivation pre-test	Distribution			
Dependent Variable	low intrinsic motivation	high intrinsic motivation	*F*	*p*
Knowledge post-test (normalized)	12	11	0,008	0,928
Knowledge increase (normalized)	8	8	0,077	0,785
Number of Exercises	15	17	0,082	0,777
Teacher´s evaluation of the exercises	15	17	1,445	0,239
Number of consumed Learning Resources	10	13	4,215	0,053
Learning Resource Evaluation	10	14	4,576	0,044
Number of weSPOT Activities	14	17	0,829	0,370

With regards to intrinsic motivation, students with a high intrinsic motivation pre-test value, evaluated the learning resources better (M = 3,70; SD = 0,55) than students with a low intrinsic motivation pre-test value (M = 3,15; SD = 0,69).

DISCUSSION

Generally speaking, the results of this study 2 are very similar as the results of study 1. Thus, for most of the results´ interpretations as well as for possible explanations of surprising or not that surprising outcomes we would like to refer to the study 1 discussion section.

Not surprisingly, all factors of the Intrinsic Motivation Inventory are highly correlated with each other. A bit more surprising is the "stability" of these relationships over time: The Intrinsic Motivation Inventory actually aims to measure a person current motivational state (i.e. motivation as a state variable) rather than a general motivational trait of a person (i.e. motivation as a trait variable). On the other hand, the "stimulus" or situation, i.e. learning with the tools on an inquiry project, remained the same and hasn´t changed over time. From this perspective one could expect at least some positive correlations between the pre-, intermediate, and post-test results.

"Intuitive" learners had lower intrinsic motivation values during and at the end of the study. "Verbal" learners were less motivated during the whole inquiry project, as indicated by lower pre-, intermediate-, and post-test values at the Intrinsic Motivation Inventory. These results could be explained to some extend when considering the topic of the student´s inquiry: the learning domain was about "Perception". Quite naturally, many learning resources consisted of pictures (e.g. of the nose, the visual system, etc.) and some learning resources were small experiments where students had to test their "perception" or "senses". This might have been more motivating for "Visual" and "Sensing" learners.

The results on the "matching hypothesis" are the same as for study 1: there is a null-correlation. Potential explanations have been already outlined above. Again, the results of the Analysis of Variance series hint at significant gender differences. As mentioned above, the results shouldn´t be over-interpreted though.

4.5. Pilot "Grazer Schulschwestern" – Cross-Studies Analysis

In this section, we analyse data which has been collected across both studies at the Graz Schulschwestern. In particular, this analysis concerns the learning style data, since the Index of Learning Style Questionnaire has been filled out by the student at the very beginning of Study 2 and at the very end of Study 2, in other words, twice within an intermediate interval of around 5 months. A correlative analysis delivers some insights on the test-retest reliability and on the stability of the underlying "personality" constructs.

In addition to that, the main outcomes of the qualitative interview with the teacher are described above. The questions of the interviewers and the answers of the interviewee are related to both studies.

Results

Descriptive Results

The following table shows the (normalized) overall values (combined for study 1 and for study 2) for the four dimensions of the Index of Learning Style Questionnaire.

Descriptive statistics (with means, standard deviations, range) of the student's learning styles, averaged over 2 measures

Variable	Mean (*Standard Deviation*)	Range: Min. empirical (*theoretical*)	Range: Max. empirical (*theoretical*)	*N*
Learning Styles: Active-Reflective	0,49 (*0,14*)	0,20 (*0,00*)	0,80 (*1,00*)	27
Learning Styles: Sensing-Intuitive	0,44 (*0,22*)	0,09 (*0,00*)	0,91 (*1,00*)	27
Learning Styles: Visual-Verbal	0,36 (*0,22*)	0,00 (*0,00*)	0,82 (*1,00*)	27
Learning Styles: Sequential-Global	0,39 (*0,15*)	0,14 (*0,00*)	0,76 (*1,00*)	27

Correlative Results

The following table encompasses the correlation coefficients between the pre- and the post-test values of the four dimensions of the Index of Learning Style Questionnaire.

Comparing the pre- and the post-test values of a particular dimensions by means of a correlation coefficient, is measurement of this dimension´s test-retest reliability.

As indicated by the first grey "diagonal", three out of four scales of the Index of Learning Questionnaire possess a high test-retest reliability: the "Sensing-Intuitive" scale, the "Visual-Verbal" scale, and the "Sequential-Global" Scale.

Correlative results (test-retest reliability coefficients) for the scales of the Index of Learning Styles

	Active-Reflective (Study 1 pre-test)	Sensing-Intuitive (Study 1 pre-test)	Visual-Verbal (Study 1 pre-test)	Sequential-Global (Study 1 pre-test)	Active-Reflective (Study 2 post-test)	Sensing-Intuitive (Study 2 post-test)	Visual-Verbal (Study 2 post-test)
Sensing-Intuitive (Study 1 pre-test)	0,009 N = 32						

Visual-Verbal (Study 1 pre-test)	0,141 N = 32	0,052 N = 32					
Sequential-Global (Study 1 pre-test)	-0,006 N = 32	0,323 N = 32	0,220 N = 32				
Active-Reflective (Study 2 post-test)	0,307 N = 27	-0,021 N = 27	0,597** N = 27	0,107 N = 27			
Sensing-Intuitive (Study 2 post-test)	-0,052 N = 27	0,750** N = 27	0,102 N = 27	0,392* N = 27	0,063 N = 28		
Visual-Verbal (Study 2 post-test)	0,329 N = 27	0,095 N = 27	0,870** N = 27	0,376 N = 27	0,604** N = 28	0,151 N = 28	
Sequential-Global (Study 2 post-test)	-0,349 N = 27	0,281 N = 27	-0,018 N = 27	0,459* N = 27	0,144 N = 27	0,487* N = 27	-0,030 N = 27

Interestingly, there are some additional relationships between the scales. The correlations between the “Visual-Verbal” and the “Active-Reflective” scales, as well as the ““Sequential-Global” and the “Sensing-Intuitive” scales in study 2 have been already mentioned. The remaining two significant relationships, again between these two pairs of scales indicate that the relationships are rather stable over time (since they compare pre- and post-test values).

Qualitative Interview with the teacher

At the 15th of August 2016, a 1.5 hour long interview with the teacher from the Schulschwestern Graz took place. Two researchers from the TUGraz were the interviewers. The interview has bene recorded with a Dictaphone. In the following, we summarised the main messages of this interview.

Interviewer: **What were your experiences with the questionnaires?**

Teacher: It was interesting for me…you read the questionnaires and you think about the underlying concepts. So you get some new insights. For me, personally, in particular the learning styles inventory was interesting and inspiring. The items are obviously the outcome of a research process – and as such, a questionnaire and its items provide a lot of hints about the underlying theory. Being engaged with the questionnaire let me think about the underlying theories, so they induced a learning process on my side and I got inspired. I mean, during my studies and my training as a teacher, for example in pedagogy, we haven´t learned that much about learning styles or motivation.

For me, as a teacher, it is also interesting to see the discrepancies between my understanding of the items and the understanding of the students. Students discussed the questionnaires… I could observe that some students discussed the questionnaires or at least some of the items – even after some time, e.g. they asked each other things like “what did you fill out? Are you more introverted or extraverted?” – just as an example. With regards to the items, maybe it is important to adapt them even more to student's level, i.e. to their vocabulary; words like "introverted" or "extraverted" were unknown to some students. Maybe it would be good idea to adapt the questionnaires together with some students.

However, of course the questionnaires require time to administer – the pre-tests and the post-tests took 1-2 school hours.

Interviewer: **What about the evaluation methodology? I.e. about the test sessions, the procedure, providing the questionnaires? Was it for you or the students overburdening to administer or to fill out this amount of questionnaires and to apply the "rules" on what to do when?**

Teacher: I think it was ok, it was neither too less nor too much. If it would have been more, I think the students wouldn´t have taken the questionnaires that serious any more. As mentioned, of course they take a while; time which is then missing for teaching. Providing something to fill out every session would be too much. I think for the next time we should consider even more the structure of the school year, i.e. the breaks and holidays – because, just as an example, we got a little bit in a hurry at the end of the summer term. With regards to my additional effort, which was required to carry out the studies as such, providing the questionnaires to the students was not a problem. However, it took much effort to support you in sorting, checking and finalizing the data which is required for you [*the TUGraz team*]. Just as an example, I received around 600 submissions from the students, such as the knowledge tests, the student's exercises, etc. I need to give feedback to the students on these issues, otherwise it wouldn´t make much sense for them, so I have to look at these submissions, I have to evaluate them, and that takes a while of course. And providing feedback to students also takes some time. I wouldn´t have had these submissions without the studies, so that was basically the additional effort on my side due to the studies.

Generally speaking, compared to my other classes, there was no teacher-centred teaching. Even if there are some disadvantages for this kind of pedagogy, the main advantage is that you principally get direct feedback from the students, if they got it or not. You can always ask them if they have some questions or they can always ask by themselves. In a self-directed, peer group learning setting, you get this kind of feedback, i.e. if and what the students grasped and what not, only once you receive and evaluate the submissions. Once you have the submissions, you can try the "see through the students eyes" to understand what has been grasped and what not. A nice example is the learning resource "structure of the eye". Usually, my way of teacher-centred teaching looks like this: I show the learning resource on the screen or make a sketch on the blackboard and go through it, together with the students, and try to initiate a discussion. That takes around half one hour and I immediately know if the students got it or not. In the self-directed learning setting, I do not have that much to do during the class sessions, however; afterwards, when I get the outcomes of their exercises, then it takes of course much more than one hour.

Interviewer: **What about the LEAs Box platform and its tools? What were your experiences?**

Teacher: What I found most interesting is that you usually don't have such a clear idea about the "mapping" between competences, the learning resources, the exercises, etc. I men, of course you have some vague ideas, but they are not explicit. When using LEAs Box, and the FCA component, where I have to fill out: what are the learning resources, what are the competences, which learning resource support which competences?, I really get a representation of my teaching project for the next weeks - I immediately see if something is missing or underrepresented. This way of thinking is new for me, I mean, of course you prepare your lessons and teaching, but so far I wasn´t thinking about also preparing a mapping between competences and learning resources. And for me that´s again a learning process for myself, I think it's helpful and important.

There is another argument: There are already new rules in the Austrian high school system in place: Teachers have to evaluate the students´ competences, as defined by the by the Austrian ministry for education [*see Bundesministerium für Bildung und Frauen, 2012*]. And if a student would have a negative overall evaluation at the end of the summer term [*which would lead to a re-examination at the Begin of the winter term*] I am only allowed to evaluate at the re-examination those competences which haven´t been mastered. And the student should get feedback on which competences he or she failed. So for these new rules, the structured way of thinking by using this FCA approach is helpful and useful. Until now it was only necessary to give a mark to the student on the general level, now it is required to give a mark or any to the different competences. It´s different and it may induce more effort; however, I think it's a good approach.

Another nice feature in LEAs Box is the opportunity that students can compare their own self-evaluations with my evaluations. Some used it intensely and also asked me why I evaluated them differently than they expected.

Interviewer: **What could be improved? What features are missing?**

Teacher: What would be nice is if students could upload their exercises directly to the LEAs Box platform, so that I can evaluate them online, mark and evaluate them directly there, and that the student receives a notification on the evaluation result. What would be also nice as a feature is a comment field where I could argue why I evaluated the student´s exercise that way. Well, not only with regards to the students' exercises but also with regards to the evaluation of their competences in the Flower app. The Flower app as such is really a nice tool and the students liked it, however, if I would have had a comment field where I briefly outline why I evaluated the student that way, it would be more convenient, because students came to me and asked me about the discrepancies between my evaluations and their self- evaluations. This is fine of course, however, it again takes time which could be used for other things. It should be more transparent for the students.

Interviewer: **How was your experience with tools?**

Teacher: Well, with regards to the configuration tool, it took some while to understand how it works, how it is structured and to become familiar with it. What seem to be missing are a documentation and a help function [*One interviewee pointed out that a documentation is available, and that there is actually a help button at the starting page at the platform. The interviewee mentioned that the next question then is if the teacher wasn´t aware of these support features or if they were insufficient or hard to understand or not useful*]. Ok, I have to admit that I wasn´t aware of that and if I had problems or questions, I directly contacted you [*the interviewees*]. However, I think there is a steep learning process: The first time it might be time consuming to define the students, the groups, the activities, etc.

But you get used to the configuration tool. And, as mentioned previously, I think this way of "thinking", in the sense of which competences are related to which learning resources and activities might take more time and effort, but at the end, there are much more advantages.

Just one suggestion on the confirmation tool: For the competences, a tree-structure view would be nice because competences have sub-competences and a tree-structure view might help you to immediately check if you structured the competences the right way. This I just a suggestion for improvement, it is not disturbing as it is now. With regards to the activities, I think they are much easier to define via the configuration tool because when defining an activity, you get the list of competences and you just need to tick the boxes. So that's pretty straightforward, simple and nice.

Regarding the MyClass tool, I also like it, but as mentioned above, a comment field would be nice so that I can argue why I evaluated a particular student´s exercise that way. Otherwise they come to me, one by one, and want to get - understandably enough - some feedback and this takes a lot of time.

Regarding the Flower app, again, I and my students liked it a lot. However, the same as for the other tools, a comment field would be helpful. Since the students came to me to get some feedback, I considered it as "work". They also should have a comment function. Maybe something like a chat functionality. Everything else is fine. You have to learn it once, and then you can profit a lot from it.

Discussion

As indicated by the correlative results, three out of four scales of the Index of Learning Questionnaire possess rather high test-retest reliabilities: the "Sensing-Intuitive" scale, the "Visual-Verbal" scale, and the "Sequential-Global" Scale. This would be an indication for learning styles as a trait variable. Proponents of different theories and models on learning styles have different opinions about the stability of learning styles (see also Deliverable D3.5: Review Article about learning styles and cognitive styles). Correlations between different scales measured at different point in time are an indication that the scales are not independent from each other. However, of course an indication is far away from being a proof, and for such a question, more participants have to be analysed.

The qualitative interview provided some interesting insights on how the teacher used the tools and the LEAs Box platform, what problem and difficulties he encountered and what he appreciated most. However, this interview only provides an overview and didn´t go to the details. During both studies, the TUGraz team was in close contact with the teacher and received many (and a great variety of) suggestions for improvements, questions, or even bug-reports. From this perspective, the teacher was much more than a "usual beta tester" in an ongoing formative evaluation process.

4.6. LPV Evaluation Study

One of the main objectives and origins of the Lea's Box project is to contribute novel, theory-driven, and in particular CbKST/FCA-based ideas to the existing poo of learning analytics methods and techniques. One of the key challenges for learning analytics, in turn, is to predict learning outcomes and student performance. This is not a trivial challenge.

An educator may want to use such predictions to say if a student will get a question correct or incorrect, or might predict if a student is proficient in a certain skill, task, knowledge component, or competence. Teachers may also build predictive models of which students need intervention to avoid failing a course. These models can then be put back into the systems in which the data was collected. A recent overview of techniques is given for example by Shahiri (2015). Initiatives such as Carnegie Mellon' s DataShop (https://pslcdatashop.web.cmu.edu/) or competitions such as the KDD Cup (http://www.kdd.org/kdd-cup) reflect the state of the art in this field.

A good portion of the existing methods are statistics-based data mining techniques. These perform well on a general, statistical basis however have clear weaknesses when operating on a level of individual learners (a good example is for example described in the use case study of Hughes and Dobbins, 2015), Also, many statistical approaches build upon a set of (at least) debateable statistical assumptions and decision criteria. A well-elaborated overview about the strength and weaknesses of practical applications of such data mining approaches has been published by Papamitsiou and Economides (2014).

In the context of the Lea's Box project we developed an approach for predicting student performance which is based on the theoretical foundations of CbKST and which might offer an interesting, top-down technique to the field of performance prediction. The approach, named Learning Performance Vector (LPV), has been described in deliverable D3.4. In essence, the idea is that the structural information about the learning domain, the atomic units of aptitude (we name them competencies), and the relationships between these competencies provide a pool of important information for predications. In addition, we can add deeper information about the individual competencies which we call "weights". These weights reflect a competency's complexity, difficulty, or importance for a domain. Together with the actual performance data of student's we hypothesized that performance predictions can be improved. In the context of the project the LPV algorithm to predict a student's Learning Horizon (LH) have been developed and implemented in the Lea's Box system.

The purpose of this study is to evaluate and perhaps validate the predictive power of the approach in relation to a simple statistical approach. We have chosen a simulation study approach because this enables us to intentionally generate the data basis and observe the algorithmic steps.

Simulating learner data

The first step for this evaluation is to simulate realistic performance data of students. To build upon a realistic data set and to be able to make comparisons between simulated and real data, we selected a data set from Carnegie Mellon's DataShop. It is a data set of "Assistments Math 2004-2005", data set id 92 (accessible at https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=92). This data set cover mathematics (which offers an easy 'playground' because it is a well-defined domain) and includes the data of 912 students. The data set is based on in total 80 competencies (knowledge components). For the simulation study we selected a subset of 11 competencies and established a competence model (prerequisite relation) among them (see the next figure). The weights are based on the inverse solution frequencies of the real data set.

ID	Competency	Weight
1	addition	0,1
2	subtraction	0,15
3	multiplication	0,27
4	division	0,4
5	fraction	0,45
6	division /w decimals	0,55
7	fraction multiplication	0,66
8	fraction division	0,7
9	fraction percents	0,8
10	fraction /w decimals	0,9

The left panel shows the select competencies and the assigned weights;
the right panel shows the established prerequisite relation.

Furthermore, we selected 12 item types and 111 items from the data set. These cover one or more of the selected competencies, partially also other competencies (as shown in the next figure). Of course, also student data is required. In the first step we simulated 15 students with different abilities (or levels of expertise). The ability parameter was defined on a scale from 1 to 10, where 1 means no knowledge in the domain and 10 means having all competencies. The parameters were simulated on the basis of a normal distribution, assuring the medium level abilities are most common and extreme position rather seldom. Finally, because this study is about prediction, we simulated 9 time points with the assumption that in the time intervals learning occurs, depending on the student abilities.

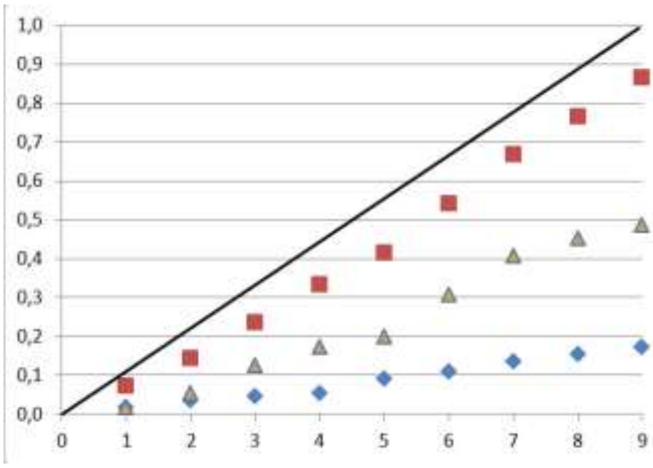
Item Type	Competencies
1	1
2	1,2
3	1,3
4	2
5	1,2,3
6	1,2,3,4
7	1,2,3,4,5
8	1,2,3,4,5,6
9	1,2,3,4,5,7
10	1,2,3,4,5,7,8,9
11	1,2,3,4,5,7,8,10
12	1,2,3,5

Student	Ability (1-10)
1	2
2	9
3	3
4	3
5	4
6	5
7	5
8	6
9	7
10	7
11	8
12	8
13	9
14	5
15	4

The left panel shows the assignment of competencies to item types;
the right panel shows the simulated distribution of student abilities.

In summary, we simulated the answer patterns of 15 students across 9 time points in 111 fictitious test items, covering 12 competencies. The simulated data set consists of 1665 data points. The following

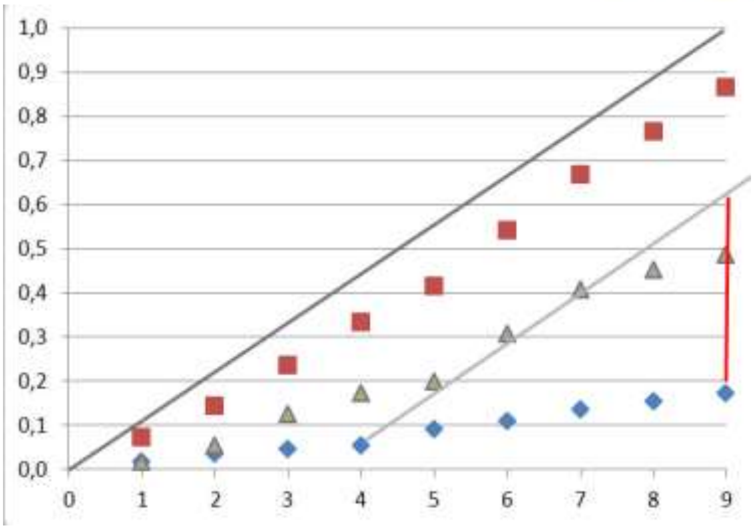
chart shows the prototypical simulated results of an excellent learner (red), a medium learner (green), and a poor learner (blue). The values show the relative increase in correctly solved items over the 9 time intervals. The bold black diagonal indicates the optimal increase, so that with each of the 9 points in time 1/9 of the items is solved correctly – or in other terms, 1/9 of the competencies have been acquired. What the results show is that the increase is determined by the student abilities, due to error rates (lucky guesses and careless errors) we see that the optimal learners is a bit below the ideal diagonal while thee poor learner still shows a slight increase. This is an expected phenomenon we can find in many data sets.



The figure shows the simulated results of three prototypical students as opposed to the ideal learning performance

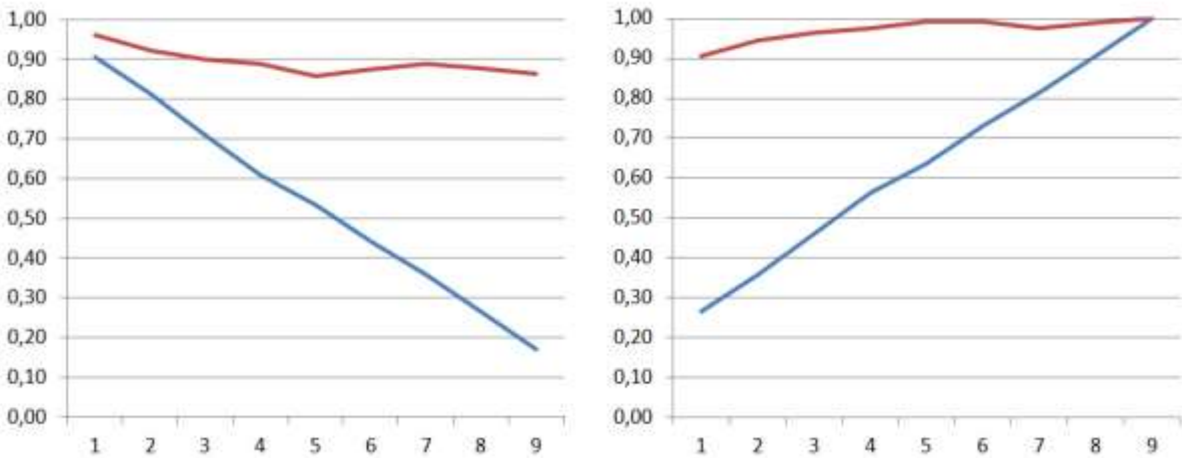
Simple statistics-based predictions

In order to evaluate and compare the quality and characteristics of the LPV, we applied a simple statistics-based idea to performance prediction, based on a retrospective view on a particular student's performance. If a student exhibits a certain performance at a certain point in time, for example the poor (blue) student in the following chart reaches a relative solution frequency of 0,05405 at the end of interval 4. One assumption that is inherent to many prediction methods is that a student might perform normally in the future. The grey diagonal that is shifted to the right indicates this idea in the following figure. This, however, is a significant overestimation of a student's abilities. There is a strong discrepancy between the final results of a student and such estimations (red line in the figure).



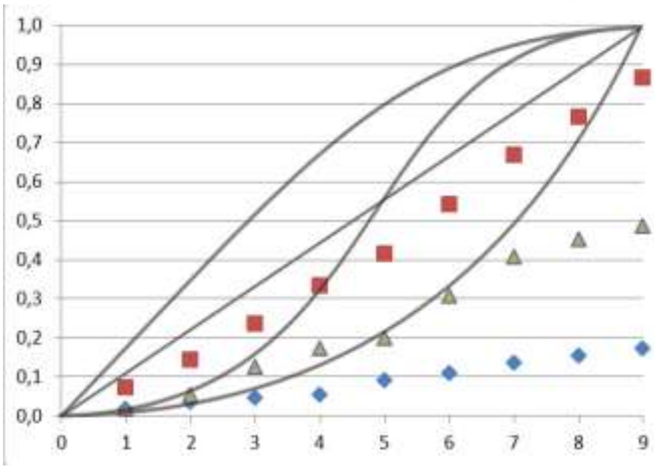
An over-simple predication approach.

The following figure shows the predictive power of this approach over time. The left panel shows the predicted end values over time for the good (red) and the poor (blue) student. The right panel shows the accuracy (difference of simulated end values and predictions) of the approach. It is evident that the method overestimates the achievements by far, even for a nearly optimally performing student. This optimal and average linear increase is a problematic approach, obviously.



The left panel shows the predicted performance over time, the right panel the method's accuracy. The red line displays an optimal student, the blue line a poor student.

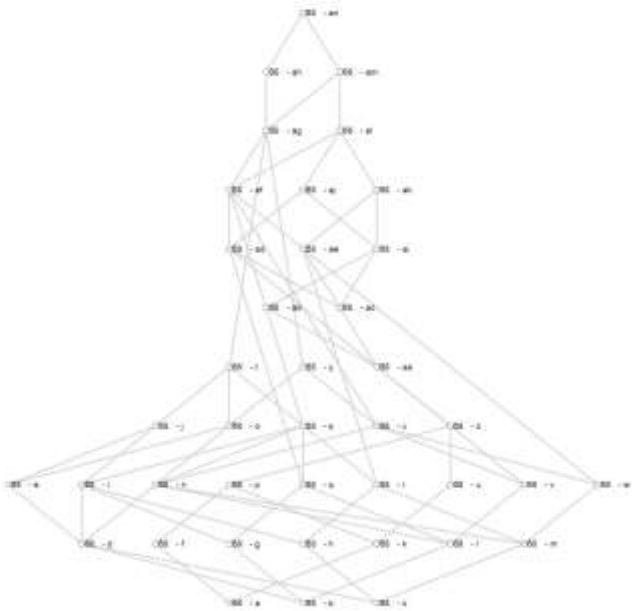
Existing approaches claim other prediction functions, e.g., various non-linear functions (as illustrated in the following figure). But also these are strong assumptions because they must find substrate in existing data. Thus, predication methods add the information of a large number of prior students to the predication model. This, however, includes again the assumption that the model is valid for all types of students (see Kim et al., 2014 for a case study).



Various non-linear predication models.

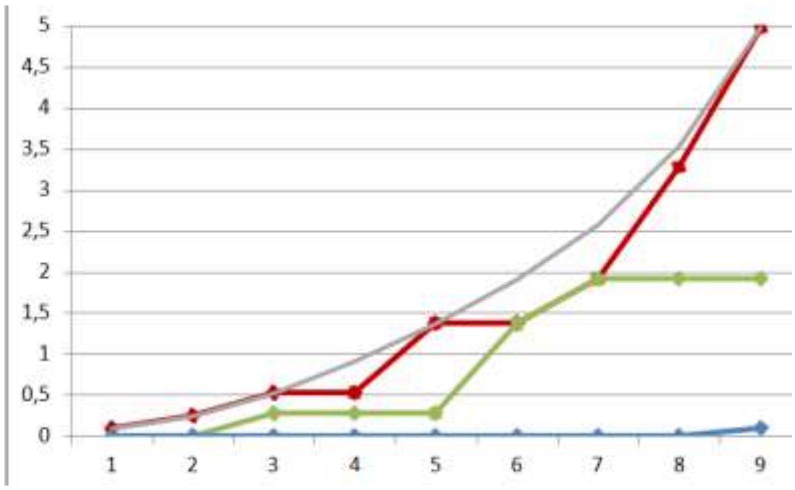
CbKST-based Prediction: The LPV

It doesn't come as a surprise that the more information are included in a prediction model, the better and more accurate it will perform. The contribution of CbKST and the Lea's Box project is to use the formal, combinatorics framework of CbKST to add information about the nature of a learning domain to the model. This information includes the number and complexity of competencies as well as the relationships between them. As opposed to one-dimensional models (see the last figure), the concept of multi-dimensional competence spaces allows for a multitude of individual learning paths and learning trajectories. The following image shows the competence space of the CbKST model introduced initially.



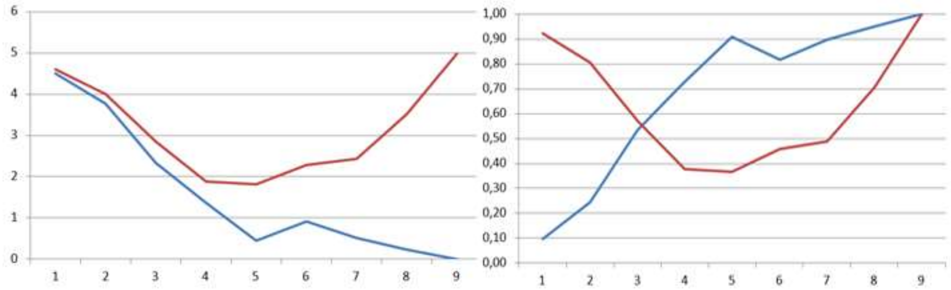
Competence space for the learning domain.

The prediction logic of the LPV is to assume a finite number of learning paths leading from the trivial competence state of having none of the competencies (the empty set) to the trivial state of having all competencies (the full set). We assume a well-graded space, claiming that in each step in the learning paths only one competency is acquired. The set of learning paths a learner is on can be identified on the basis of the current and past answer patterns (i.e., which item types having been mastered and which not). The various paths can be characterized by their complexity, which is determined by the weight of the individual steps. Thus, the performance of a learner at a pint t is characterized by the weights this student mastered up to the current point in time. This means that the order of learning specific competencies and also the order of the assessment do not distort the prediction. In other words, mastering a lot of low complexity items at an early stage is a weak indicator because major challenges are still ahead for the student. In turn, mastering highly complex items (with high weights and perhaps a larger number of prerequisites) is a very strong indicator because all prerequisite items are assumed to be possessed by the learner. The following figure shows the prediction results for the same simulated data set and the same students.



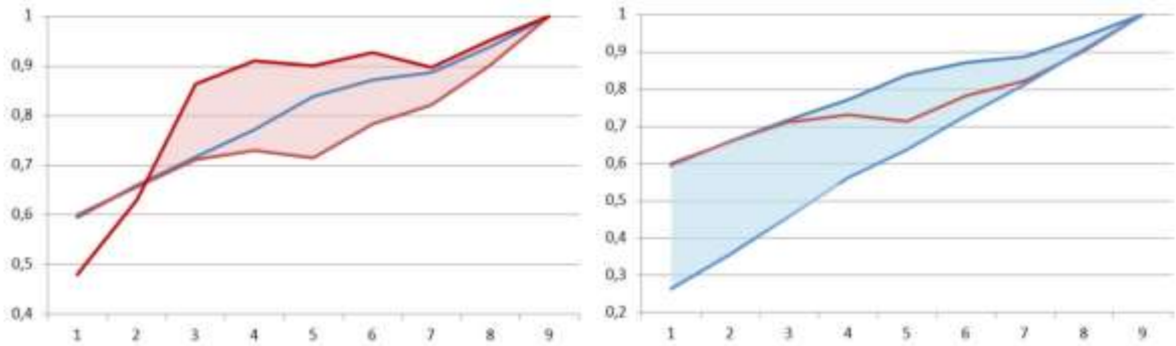
Predication results of the CbKST approach for a good (red), medium (green) and poor (blue) student.

In this example, the sum of weights assigned to the competence structure is 4.98 (the grey curve in the figure indicates the average prediction model of this approach, contrasting the linear approach we described above). The curves show the progress - in terms of mastered weights – of the three prototypical students across the 9 time intervals. The following figure shows the predicted end values at each point in time. We see that at first the predications are too high for the poor student, however, they rapidly decrease until the 5th point in time. Form here, the algorithm quickly approaches the final simulated values. The right panel of this figure shows the prediction accuracy of the final value over the 9 points in time. The result shows that the prediction with the CbKST approach is weaker for an optimally performing student as opposed to a weakly performing student. In other words, the linear approach loses predictive power with an increasing deviation from the ideal linear, diagonal increase.



The left panel shows the predicted performance over time, the right panel the method's accuracy. The red line displays an optimal student, the blue line a poor student.

When simulating different answer patterns in the 12 item types, we found that the order of item types strongly influences the power of the LPV approach. If more difficult items are presented already at an early stage, which means that the additional structural information influence the predictions, highly accurate predictions occur fast, in our example already after time interval 3 (left panel of next figure). In general, the LPV offers effective and accurate predictions. The linear approach, as said, is strongly dependent on the deviation of learning patters from the diagonal. For weak learners, for example, the predication is quit inaccurate (right panel of next figure).



Comparison of predictive power (LPV vs. linear)

Discussion

The aim of this simulation study was to find systematic evidence whether and to what extent the LPV is a suitable method to predict students' performance. Also, simulation studies allow us to explore the characteristics and dependencies of the method in its application to various data characteristics.

By the benchmark of the linear method we found stable and promising results. Specifically since the linear increase was the conceptual basis of the simulation algorithm as well.

Another critical aspect is the weighting process for competencies and subsequently assessment items. This has a strong influence on the predictions and can be based on several approaches. The simplest is a manual assignment of weights by teachers. This, however, bears the peril of an arbitrary and unfounded weighting. The strength of this approach could be that the weights a grounded on the very concrete and practical experiences of a teacher. A second and more data driven approach is to refer to the solution frequencies of items in large data sets. This is the method we used in this study. If items are solved with a high frequency, we can assume a low difficult of the competencies covered by the item and also a low predictive power in terms of CbKST-type prerequisites between the competencies. A third method we must explore in future steps is the so-called Component Attribute Approach (Albert & Held, 1999). This theoretical approach describes test items (problems) by components and their attributes. Components are major characteristics, for example which algebraic operations are included in a math item. The attributes describe the individual components, for example stating which types of numbers (positive integers, decimals, …) are part of the problem. It could be shown that the prerequisite relation of the problem space could be derived by the set inclusion principle. In our context decomposing and analysing the components and their attributes can support the weighting process. Finally a forth method is to analyse the items on the basis of their cognitive depth. This refers back to the famous taxonomy of Benjamin Bloom, revised by Anderson & Krathwohl (cf. Anderson, 2013). In the so-called Concept – Action Verb approach (Heller, Steiner, Hockemeyer, & Albert, 2006), a competency is defined by a concept, ideally in form of a proposition (e.g., "house – has – window") and a specific cognitive depth that ranges from mere "knowledge" to the level of "creation". Bloom proposed 6 such levels. An example would be "understand that a house has windows and apply this understanding in a new situation". The taxonomy also separates the knowledge dimensions factual, conceptual, procedural, and metacognitive knowledge, which in the end established a 2-dimensional hierarchy. In our context this taxonomy provides a scaffolding to analyse the items, to identify the covered competencies, and to rank the competencies according the taxonomy – which I the end specifies the weight.

Future steps will continue to explore the characteristics of the LPV under various context conditions. Also more in-depth studies using the real data sets from the "Assistments Math 2004-2005" data set.

MAIN CONCLUSIONS

- The LPV is a suitable method to predict learning performance.
- The LPV has particular strength for rather medium and poor students and it works best when at each assessment point in time a broad spectrum of item difficulties are presented.

5. PILOTING AND EVALUATION IN TURKEY

The following table summarizes the plan (left) and the actual activities (right) carried out in year 3 in Turkey, as foreseen in DoW and later by the D1.2 Annex. Evaluation activities always directly followed pilot activities, however there are evaluation activities towards exploitation rather than quality or acceptance studies of the outcomes. The actual activities are hashtaged to be tracked easier through the text.

Piloting	
Speedreading Use Case May 2016 – June 2016 100 learners, 10 teachers	#SEBIT_Co-DesignStudy_1: With Maya School Network March 2016 8 teachers #SEBIT_PilotStudy_1: With Ayse Abla School Speedreading Use Case May 2016 – June 2016 133 learners, 10 teachers
Learning Spaces May 2016 – June 2016 100+ learners, N/A teachers	#SEBIT_PilotStudy_3: With Adaptive Curriculum (106 students – Feb 2015) and/or RAUNT data (4657 students – June 2016) on Learning Spaces Tools evaluated by 100 head teachers
	#SEBIT_PilotStudy_2: With xAPI
Evaluation	
TAM3 Evaluation of private school target audience June 2016 100+ learners, 10+ teachers	#SEBIT_EvalStudy_1: With Ayse Abla School May 2016 – June 2016 133 learners, 10 teachers
Structured Interview Evaluation of private school target audience June 2016 100+ learners, 10+ teachers	#SEBIT_EvalStudy_1: With Ayse Abla School May 2016 – June 2016 10 teachers, 2 headmasters, 1 school owner #SEBIT_EvalStudy_6: With Bahçeşehir School Network 10 teachers, 2 headmasters, 1 school owner #SEBIT_EvalStudy_7: With Vizyon School Network 12 teachers, 3 headmasters, 1 school owner
Structured Interview Evaluation of "sponsors" target audience October 2016 – November 2016 100+ learners, 10+ teachers	#SEBIT_EvalStudy_2: With Abdullah Gül University December 2016 Rector and Vice Recors #SEBIT_EvalStudy_3: With Turkish MoNE December 2016 100 head teachers
Survey Evaluation of "retail" target audience October 2016 – November 2016 100+ learners, N/A teachers	#SEBIT_EvalStudy_4: With Innova A.Ş. #SEBIT_EvalStudy_5: With IBM USA

Some of these studies are for evaluating the feasibility of specific business cases; some larger studies are for evaluating the Lea's Box system as a whole in live action; but ALL of the SEBIT studies this year were geared towards exploitation.

Y3 of the project started in March 2016. It was right after the last pilot study of Y2 with Maya School Network in February 2016. That study involved a commercial product from SEBIT which had Lea's Box tool as an add-on service for real-time reporting of analytics results to students and teachers. During the 2 weeks' study students followed proposed course plan and planned their exercises based on OLM reports. All the work done by students was off-class hours. The results of the study were delivered in detail in D5.5 and though the benefits were absolutely clear, it pointed to a few potential ways to have even more impact.

5.1. Co-Design Studies

#SEBIT_Co-DesignStudy_1 was done with 10 Maya School Network teachers to chart out how to have more impact in using Lea's Box tools for planning and online course taking.

The outcome of this co-design study was the following recommendations:

1 – Students must be trained in advance about what are competencies and skills, how they are different then knowledge sets. Without such understanding it is not easy for them to make meaning out of LEA's competency state diagrams.

2 – There is a need for more scaffolding as to how best the analytics reports can be used during the early days of the study.

3 – For the students to get used to the OLM diagrams they should observe them alongside the course material and monitor their situation continuously so that they can reason about the outcomes of their performances.

5.2. SEBIT Design Study

The first pilot study of Y3 was designed based on these recommendations. #SEBIT_PilotStudy_1 took place in Ayse Abla School in June 2016 and lasted 4 weeks including the training period and 12 days speedreading course assisted by Lea's Box OLM tools. 133 students participated from 8 different classes of 6th and 7th grades. They spent an hour every day in the computer lab, doing up to 30 minutes of speedreading structured course activities. One half had OLM diagrams open and updated in real-time alongside the course activities, and other half examined them after the course activities. All groups had negotiation sessions every other day. The inquiry lines of the study were

1 – How effective would analytics reports be if they were used in real-time alongside the digital learning activity?

2 – Which uses of OLM charts lead to a negotiation/persuasion session, initiated by the student?

3 – Would longer a priori training help being more mindful about one's skills during a period of training?

4 - What should be the target age group if Lea's Box was incorporated into HızlıGo speedreading product as the first market implementation of the project? To answer this question, 6th and 7th grades were taken to the pilot, because grades 9 to 11 were already piloted in March 2016.

5 - What should be the strategic positioning in school study cycle? To answer this question, the pilot was carried out in the school computer lab, because the course was assigned as a homework, already in March pilot?

6 - What should be the convention in using analytics results? At will, whenever the students feel the need to evaluate his/her progress and plan his/her training OR regularly, during each study session as part of the course plan? To answer this question, the two conventions were tried at the two grades. 6th grade students used Lea's Box routinely as part of their course plan, while 7th grade students 1 week after their course started and used Lea's Box in special, dedicated sessions.

Summary of results

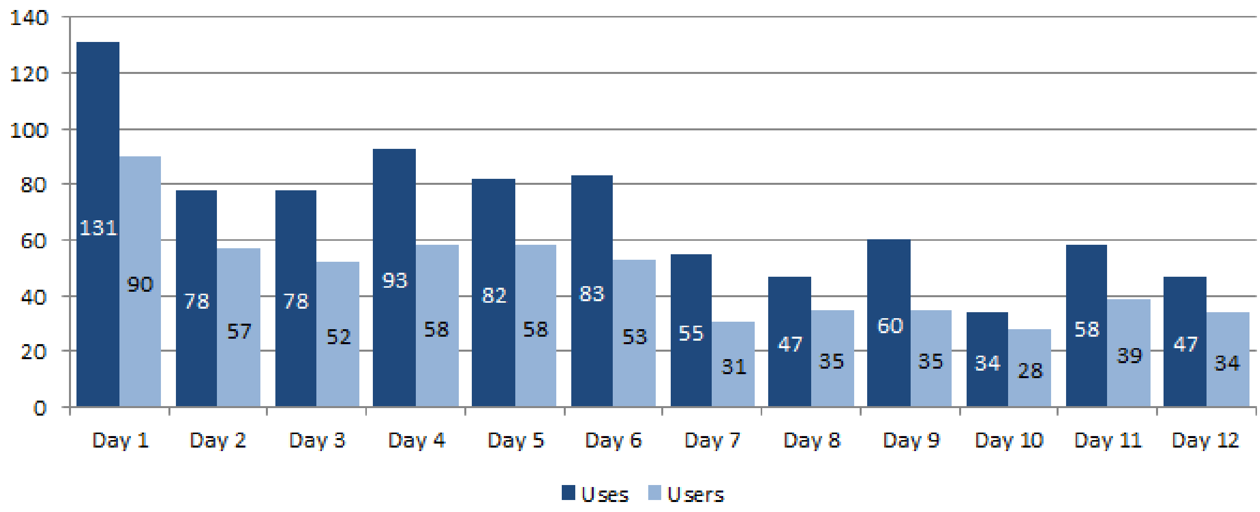
The results of this study is documented and submitted to a number of journals and conferences.*As a summary, the students had benefited more from this course more than their peers in Maya school. Specific findings were that the study showed that the most used visualisation were the level of activity, the skill meters and the table, but the across time visualisation was more likely to lead to persuasion episodes. Otherwise, students seem to be interested in accessing various aspects of learning analytics, and not only to the current state of their model, like the number of activities performed each day and the evolution of their model during the course. An interesting result was that as the students learned about what study patterns worked for them, their use of the analytics reports faded.*

The below table presents the uses of Hizligo speedreading digital course and the OLM by the 133 students and their final level, i.e. the level of their model at the end of the course.

Uses of Hizligo and the OLM.

	Final level	Activities performed	OLM sessions	Views of the model
Nb. students involved	133	133	128	128
Sum	-	49149	855	3282
Range	44 - 100	122 - 648	1 - 46	1 - 160
Average	65.80	369.54	6.68	25.64
Median	64.49	382.00	6.00	19.50

In average, the final level of the students is 65.80, after in average 369.54 activities performed. That correspond to a very good level of involvement in the course as if a student follows only the recommendation of activities for every day of the course without doing additional activities, (s)he would do 193 activities during all the course. The figure below shows that as the students became more familiar with OLM, and learned about their study patterns the use of analytics has faded.



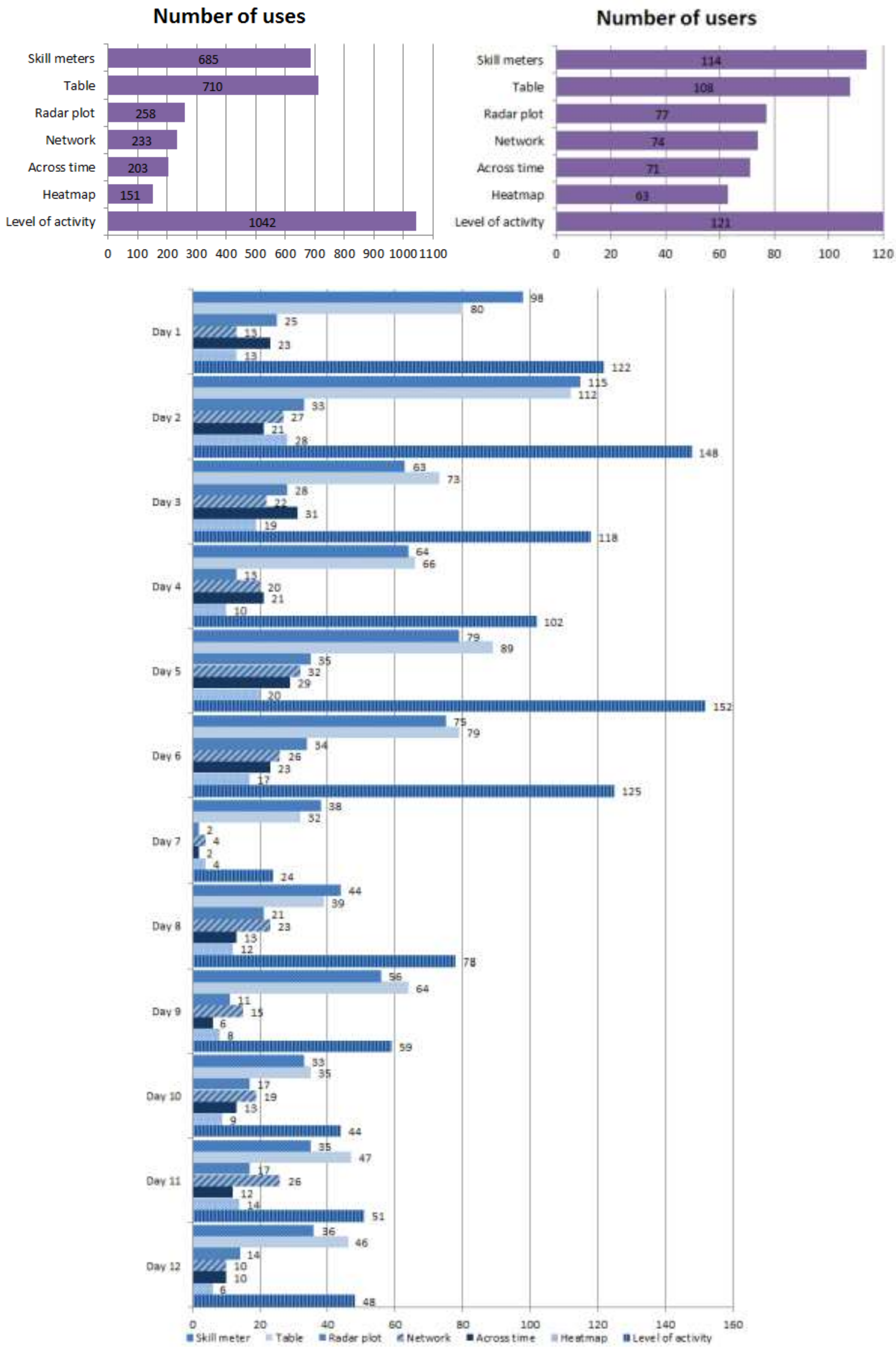
Use of the OLM during the 12 days' course.

The below table presents the uses of each visualisation. We can see that the visualisation with the most users and uses is level of activity, followed by skill meters and table.

Use of each visualisation.

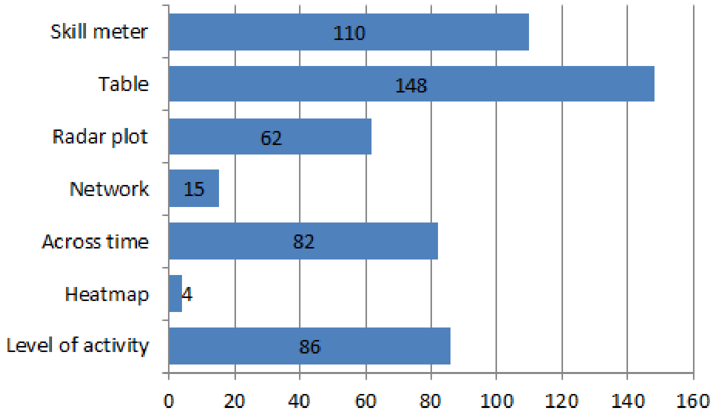
	Nb. users	Nb. uses	Range	Average	Median
Skill meters	114	685	1 - 39	6.01	4.00
Table	108	710	1 - 42	6.57	5.00
Radar plot	77	258	1 - 13	3.35	2.00
Network	74	233	1 - 11	3.15	2.00
Across time	71	203	1 - 12	2.86	2.00
Heatmap	63	151	1 - 10	2.40	2.00
Level of activity	121	1042	1 - 47	8.61	5.00

This tendency is confirmed by the use of each visualisation during the 12-day course. There is no significant change of use of the different visualisations over the course.



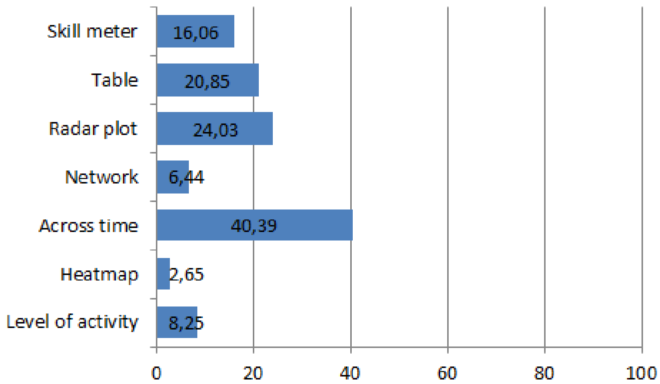
Use of each visualisation during the 12 days' course.

During the course, 507 persuasion episodes have been initiated by 78 students (58,65% of the participants). 326 persuasion episodes were resolved (64,30% of the persuasion episodes), that to say resulted in an updated of the model after an agreement between the student and the system, however, in 181 episodes, no agreement was found between the system and the student so the persuasion remained unresolved and the model was not updated. The below chart presents the number of views of the model that lead to persuasion, for each visualisation.



Number of views that lead to persuasion for visualisations.

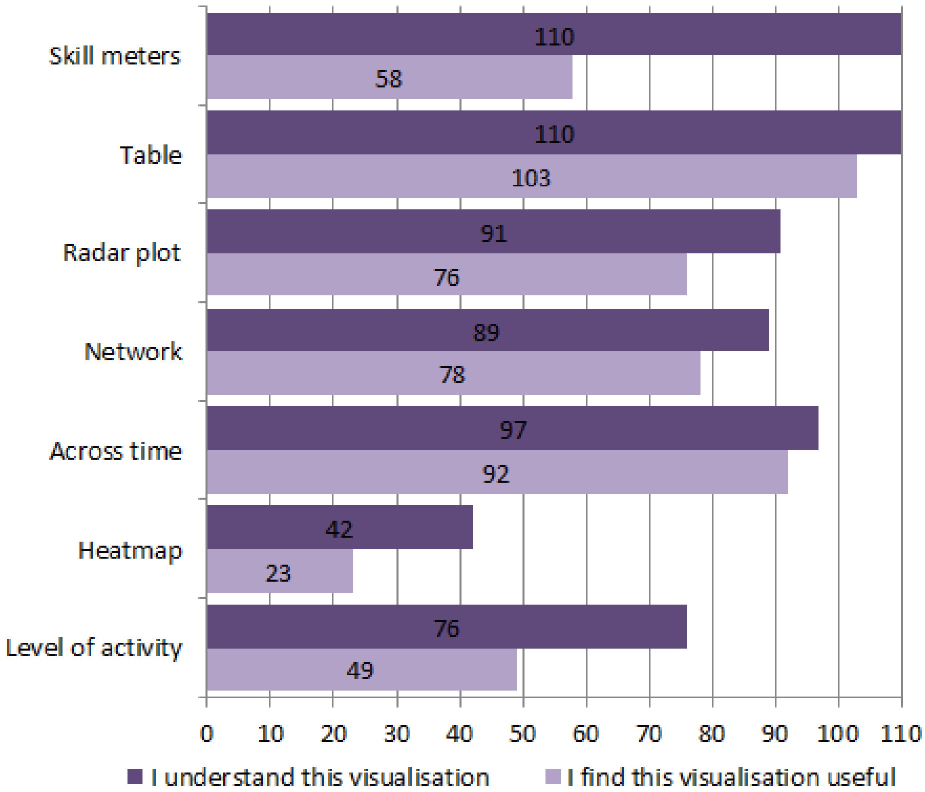
We can see that table and skill meters are the visualisations that mainly lead to persuasion: table lead to 148 persuasions (29.19% of the persuasions), skill meters lead to 110 persuasions (21.70% of the persuasions). The level of activity visualisation lead to 86 persuasions (16.96% of the persuasions), across time lead to 82 persuasions (16.17% of the persuasions) and radar plot lead to 62 persuasions (12.29% of the persuasions). Two visualisations were not much used before persuasions episodes: network lead to 15 persuasions (2.96% of the persuasions) and heatmap lead to 4 persuasions (0.79% of the persuasions). As each visualisation have not been used the same number of times, the below chart represents the percentage of views of each visualisations that lead to persuasion.



Percentage of views that lead to persuasion for each visualisation.

We can see that if the level of activity visualisation was the most used to view the model, only 8.25% of the views lead to persuasion. At the contrary, the across time visualisation is in proportion the visualisation that lead to the most persuasion, in 40.39% of the case, a student that viewed his/her model with the across time visualisation initiated a persuasion.

Below are the results of the final questionnaire completed by all the students and relative to the perceived understanding and usefulness of each visualisation.



Final questionnaire: perceived understanding and usefulness of each visualisation.

According to students, the easiest visualisations to understand are skill meters and table, considered as understandable by 110 students (82.70% of the participants), followed by across time (97 students, 72.93% of the participants), radar plot (91 students, 68.82% of the participants), the network (89 students, 66.92% of the participants) and the level of activity (76 students, 57.14% of the participants) and the heatmap (42 students, 31.58% of the participants). According to students, the most useful visualisation is table (103 students, 77.44% of the participants and 93.64% of the students that claimed to understand it), followed by the across time (92 students, 69.17% of the participants and 94.85% of the students that claimed to understand it), the network (78 students, 58.65% of the participants and 87.64% of the students that claimed to understand it), the radar plot (76 students, 57.14% of the participants and 83.52% of the students that claimed to understand it), the skill meters

(58 students, 43.61% of the participants and 52.73% of the students that claimed to understand it), the level of activity (49 students, 36.84% of the participants and 64.47% of the students that claimed to understand it) and the heatmap (23 students, 17.29% of the participants and 54.76% of the students that claimed to understand it).

The statistical analysis given above concludes that the most used visualisations are not necessarily the most likely to lead to persuasion. Indeed, it seems that the simplest structured visualisations, the across time, table and skill meters, are more likely to lead to persuasion. The most difficult visualisation according to these secondary-school students was heat map, a visualisation firstly intended for teachers. However, we notice that it has been used and is found useful by some students despite their age. What's more, students seem to be interested in accessing various aspects of learning analytics, and not only to the current state of their model. Indeed, the level of activity and across time visualisations have been used a lot and presents information about the number of activities performed each day and the evolution of their model during the course.

5.3 Evaluation Study at Ayse Abla School

#SEBIT_EvalStudy_1: With Ayse Abla School. Following the pilot study an evaluation study was carried out using TAM3 framework (see Appendix). Survey Monkey was used to collect input from all the students and the factor analysis of the variables that load "perceived usefulness" and "perceived ease of use" for Lea's Box were calculated.

Notice that a similar evaluation study was carried out after the Y2 pilot study (reported in D5.5 Piloting and Evaluation Report 2 - March 2016) which revealed that both students and teachers have a strong behavioral intention to use Lea's Box OLM tools, but those intentions were reduced by mostly "perceived usefulness" factors in case of teachers, and mostly "perceived ease of use" factors in case of students. It was clear that usability factors had to be improved which obviously effect the exploitation plans as well. The major reason for usability drawback was the commercial product used and LEA's Box were two separate sites, with separate login and navigation paradigms, practically doubling the cognitive and time burdens of use.

The end result of the TAM3 evaluation is as follows:

1. Behavioral intentions towards using Lea's Box analytics (at least in the context of speedreading skills development) decline about 10 base points on the average as the grades get lower. On the other hand, when used simultaneously with the learning product, perceived ease of use increase substantially (about 5 base points). **In conclusion, the product should rather be positioned for high school students and the analytics models must always be present on display, with real-time updates. Therefore, the features must be integral to the product**.

Factor	High School Teachers (%)	High School Students (%)	Middle School Students (%)
Perceived Usefulness	88	80	71
Perceived Ease of Use	83	88	77

2. Using analytics as part of a course cycle dramatically increases the usage time, but strong mentoring becomes essential as students are affected very much from each other's comments on the analytics results. The correlation between responses within one classroom was almost double the value correlation between different classrooms. **In conclusion, the product would require strong mentoring and role modeling by the teachers to be used consciously and effectively.**

Relying on these evaluation studies the value-added speedreading software is planned to be release in January 2016, using the final release of the Lea's Box tool (delivered later in October 2016), relying on the signing of an MoU and open source licensing of the software. In case of success, we would achieve the first **MARKET IMPLEMENTATION** of the outcomes in the project lifetime.

5.4. EVALUATION OF INTERACTIONS WITH SPONSORS IN TURKEY

Sponsors includes entities that have authority to make a bulk purchase for a large set of end-users. Their main business requirement is robust scalability and standards based data interoperability. After having tested Lea's Box tools, in particular OLM in live trials that lasted weeks, SEBIT run a meeting with Turkish MoNE to introduce the project outcomes. SEBIT is delivering Turkish MoNE a national digital educational content moderation and reutilization platform. Lea's Box has the potential to serve as a data infrastructure for that platform.

#SEBIT_EvalStudy_3: TR MoNE Recruits 100 Head Teachers

As a result, 100 head teachers recruited by MoNE was presented with Lea's Box system as a whole, trained and tried out the system for #SEBIT_EvalStudy_3.

The whole system evaluation study with 100 head teachers took place in Ankara in December 2016. These teachers already had a training on the national content moderation and reutilization platform called EBA. EBA has in store a large set of digital learning events which can be instructional content as well as assessment items. The positioning for Lea's Box was that the data accumulated in this platform could be utilized using analytics tools for a number of **pedagogical problems given in the annex.**

After a guided and then free indulgence with Lea's Box tools a survey was made on how likely would Lea's Box be of use on a Likert scale of 5. Rounded average responses were as follows:

	PEDAGOGICAL PROBLEM	Lea's Box WOULD BE A SOLUTION
30.	How to create accurate personalized assessment relevant to each student's progress of learning content?	AGREE
31.	How to assess progress? How to assess value added by school/teacher?	STRONGLY AGREE
32.	How can students/teachers meaningfully reflect on student work to create higher quality work?	NEUTRAL
33.	How does information disseminated by peers as well as peer-to-peer learning effect knowledge acquisition?	DISAGREE
34.	How do assessments promote student understanding and growth?	AGREE
35.	Which assessments measure best?	NEUTRAL
36.	How to find an ideal balance between the urge to gather assessment data and having sufficient time for teaching? What can be omitted and what definitely needs to be recorded?	STRONGLY AGREE
37.	How to vary instructional strategies to increase learning efficiency?	NEUTRAL
38.	What assessment alternatives can be incorporated in order not to focus solely on grades?	NEUTRAL

39. How can efficiency of new teaching/learning models be assessed?	NEUTRAL
40. How can teachers save time while creating individualized development/learning plans?	AGREE
41. Who are the best/good and who are the low performers?	STRONGLY AGREE
42. What to teach next - /wrt to the entire class and /wrt to individual students?	STRONGLY AGREE
43. How to build groups among students (e.g., for projects, presentations, homework, etc.) – which students match which don't?	NEUTRAL
44. What materials / topics should individual students attend – what homework would be the most effective for an individual?	STRONGLY AGREE
45. What do have students in common, where are they different?	AGREE
46. Which social standing do students have in a group?	NEUTRAL
47. What test items and evaluation task would be optimal for a particular student?	AGREE
48. What are suitable learning paths?	AGREE
49. How is a domain structured?	AGREE
50. What are the differences between students' perception of a domain in comparison to the teachers'?	NEUTRAL
51. How to evaluate one's own teaching or specific educational projects?	AGREE
52. Is the teaching material appropriate?	STRONGLY AGREE
53. Do students reach the goal of the course?	AGREE
54. Is there a risk for particular students to fail in a course?	NEUTRAL
55. How to address cross-subject, cross-domain meta-skills within a class or across the entire school?	NEUTRAL
56. How to give tailored, understandable, and formative feedback – students – to parents – how to inform colleagues?	AGREE
57. How to select suitable learning materials / modalities for individual students (in terms of learning styles, for example)?	DISAGREE
58. How to profit from all the various technologies that are available (apps, games, online tools, etc.) best?	DISAGREE

5.5 Prospects for the system as a whole in Turkey and in Europe

The studies described in this section point to a real potential for Lea's Box system to be used in Turkey, not perhaps a whole but tailored towards specific use cases. The analysis of both usage data and the survey results of many studies committed prove that Lea's Box competency based assessment tools can be definite value adds in 3rd party digital learning platforms, such as those of SEBIT. Not every tool in the box seems to be ready for prime time, but a big sponsor such as MoNE can choose to keep the whole box if the most relevant tools are tailored to the national digital learning platform. For a more definitive discussion The SWOT analysis in D6.5 Dissemination and Exploitation Plan and the strategic positioning that follows can be referred.

5.6 Evaluation of the Specific Business Cases in Turkey

Each Business Case was evaluated in Turkish studies. Below is an accounting of these studies in brief with raw outcomes:

#SEBIT_EvalStudy_6 of Business Case 1: Value-Added Feature in a SEBIT product:

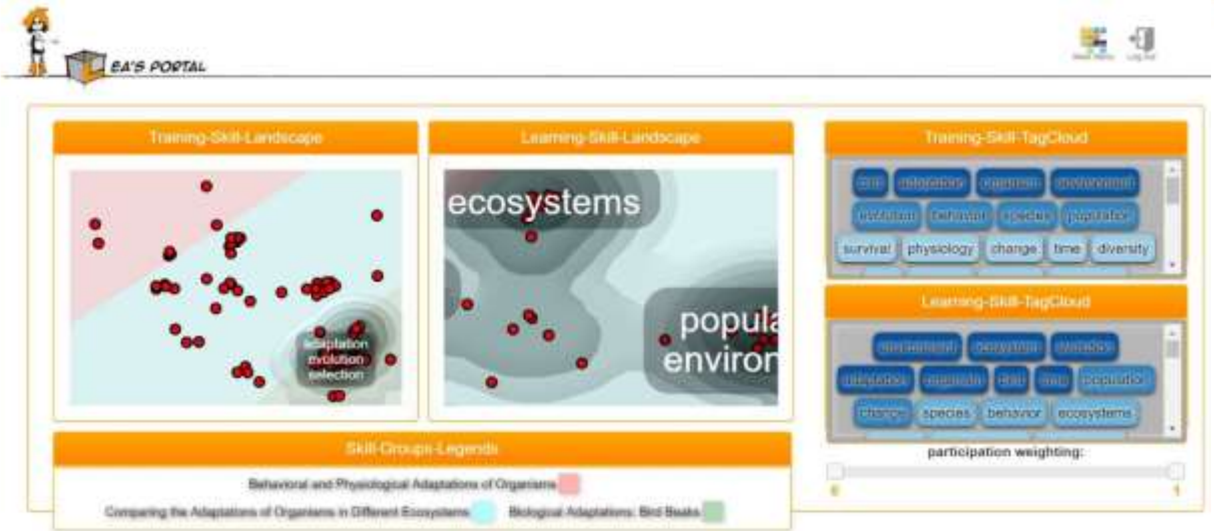
This study was carried out with Bahçeşehir School Network which has some of their schools who are HızlıGo speedreading application users. With mock OLM charts embedded in the product a dedicated meeting was held with decision makers. Their unanimous verdict was a 2.0 version of the product with such reporting features had an increased chance of order renewal.

#SEBIT_EvalStudy_7 of Business Case 1: Value-Added Feature in a SEBIT product:

This study was carried out with Vizyon School Network which has all of their schools using HızlıGo speedreading application. In fact, this school is a design contributor to the product. With mock OLM charts embedded in the product a dedicated meeting was held with their instructional designers and decision makers. Their unanimous verdict was a 2.0 version of the product with such reporting features had an increased chance of order renewal.

The other indicated products by SEBIT which can have features from Lea's Box are Adaptive Curriculum and RAUNT. Developed by SEBIT, RAUNT is the first and only customized university exam preparation platform in Turkey. Partnerships were established with publishers and RAUNT was launched for the individual subscriptions customers during 2015-2016 academic year, followed by private school corporate customers during 2016-2017 school year.

In **#SEBIT_PilotStudy_3: With Adaptive Curriculum and/or RAUNT products**, a data set from both products have been provided to the research partners to be tested with Lea's Box tools. Below is the screenshot of the Landscape tool in Lea's Box, which processed real data from SEBIT US product Adaptive Curriculum. More of such trials with new tools in System Final Release (D2.6 – October 2016) are being carried out as of the writing of this deliverable.



The Landscape tool as well as others in the final system released were evaluated at **#SEBIT_EvalStudy_3** with 100 head teachers on their 2nd day of Lea's Box workshop.

Following the introduction and demos a survey was carried out to assess the Technological Readiness Level of these tools[3].

[3] EARTO (2014), The TRL Scale as a Research & Innovation Policy Tool, EARTO Recommendations, retrieved from: http://www.earto.eu/fileadmin/content/03_Publications/The_TRL_Scale_as_a_R_I_Policy_Tool_-_EARTO_Recommendations_-_Final.pdf

All the tools were regarded to be at TRL 2 Interaction with Users, Designers and Engineers, except for the Landscapes tool, which was regarded to be at TRL3 Exploring Market Opportunities stage of maturity. Therefore, it deserves to be followed up as a future opportunity in learning analytics.

#SEBIT_EvalStudy_4 of Business Case 2: Value-Added Feature for a 3rd Part Vendor:

This study was carried out with Innova AŞ. which is one of the largest software houses in Turkey with nearly 1000 programmers on payroll and Innova is a sister company to SEBIT, both of which are owned by Turkish Telecom Group. A dedicated meeting was held with Innova officers to discuss if Lea's Box system or any features therein could be an add on for their products or turn-key projects. The outcome recommendations were as follows:

1 – All and any license issues has to be cleared out

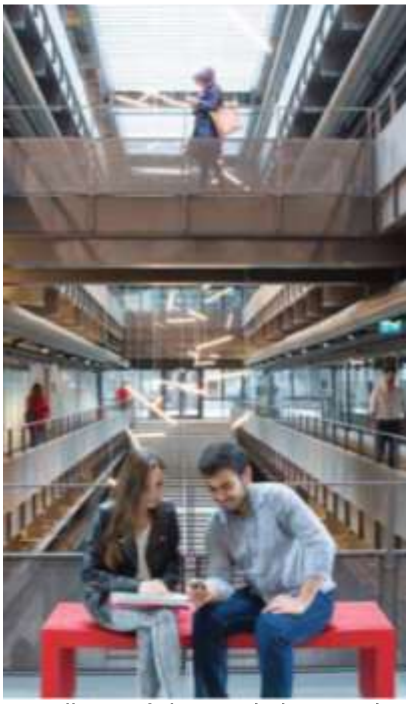
2 – Scalability testing has to be performed to at least a thousand simultaneous users or beyond.

2 – The interface should rather be a standard (IMS Caliper or ADL xAPI)

Of the last item on the list a study is still being done involving TUGraz and SEBIT. **#SEBIT_PilotStudy_2** has the same scope as with other pilot studies SEBIT has committed, except the interface to the Lea's Box is xAPI, rather than a proprietary adaptor. An xAPI interface has already been released for Lea's Box (see D2.6 Final System Release, Section 7). An xAPI recipe is being prepared so that it can be used with SEBIT products.

#SEBIT_EvalStudy_3 of Business Case 4: Value-Added Feature in a National Solution:

#SEBIT_EvalStudy_3 was a hands on 2 days' workshop with 100 head teachers recruited by the Turkish MoNE. These teachers are likely early adopters for a potential integration to the national digital learning platform EBA. Various survey results and recommendations out of this study have already been reported above. To summarize Lea's Box platform has the potential to be integrated to EBA which is developed by SEBIT and will be updated by SEBIT with new functionalities during the next 8 years to come. However, in light of the new data privacy law in Turkey, further studies have to be made to ensure the requirements of the law are met (such as stateless encrypted tokens that hides the user's identity even for the data processing entity), data processing is democratized using a standard (eg. xAPI) and scalability to 20+ Million users has to be tested.



#SEBIT_EvalStudy_2 of Business Case 5: Learning Analytics for HEIs

This business case has been tested with Abdullah Gül University (AGÜ) in Kayseri which is one of the new but technically advanced universities in central Anatolia. One specific use case which is very much on demand in HEI is academic counselling based on evidences revealed by learning analytics. AGÜ is aiming to launch competency based degree programmes. In such universities, learning analytics can be employed to facilitate “personalization of the competency development”. Unlike intelligent tutoring software which tries to automate personalization by adapting course content, learning analytics enables a human-in-the-loop approach in which instructors or the students themselves can differentiate content and/or assignments based on their readings of the analytics results. Analytics can also be used to closely monitor the engagement and progress of sub-groups of students, such as those who follow the course as part of their minor studies and so target a less demanding level of competency achievement.

Students in higher education need more insight when deciding about and following their studies. AGÜ offers a large set of educational opportunities and projects. Students’ decisions about their transcript portfolio, major and minor subject areas determine their grad profiles. Therefore, they seek support via academic counselling. The curricula of AGÜ academic programmes vary significantly for each student. The abundance of course offerings from faculties and departments, undergraduate research opportunities and service learning activities present a multitude of directions and transcript portfolio compositions. The path through the curriculum is always a trade-off between the abilities of the students and job market expectations upon graduation. SEBIT is designing an e-platform specific for this kind of HEI. Lea’s Box analytics tools can be configured to work in conjunction with this e-platform to discover levels of engagement per course or course module, ensuing success level, and overall learning outcomes for the qualification. These analytics results can be used to compare the student trajectory with alternative paths and help decide on the target qualification (an example is available at Open Universities Australia, Personalised Adaptive Study Success initiative). Finally, shifting the focus to faculty, it is fast becoming a common practice of the teaching staff to use data to improve their own practice and many institutions proactively use learning analytics as a diagnostic tool at a systematic level (e.g. informing the design of modules and degree programmes). In pursuit of quality assurance and quality improvement it is even possible to create new metrics of quality, following of which leads to more likely, even guaranteed graduates of targetted proficiency.

The case of AGÜ is particularly valuable since it is supported by the Ministry of Development and expected to set an example to other HEIs in Turkey.

A workshop with above context was planned to be held in AGÜ, in December, but unfortunately it couldn’t be achieved due to busy schedules of attendees in this month. The event is postponed but a dedicated meeting is set up with responsible vice rectors and the rector on 28th December. As of writing this deliverable, the opportunity with HEIs is still quite open.

#SEBIT_EvalStudy_5 of Business Case 5: Learning Analytics for Content Evaluation Services for Publishers

SEBIT is in a working partnership with IBM Watson Services for Education. This study started with a horizon seeking event held via telco with IBM USA unit in business development for IBM Watson Services for Education. Lea's Box can leverage its CbKST features to see exactly how much a given content product impacts learning outcomes. This evaluation can provide feedback about the validity of the knowledge structure and about the relative difficulty of the assessment items. Publishers can use this service to update items or content parts. However, sifting through a large set of content assets is time consuming and tedious. IBM Watson has very advanced natural language processing abilities which can dramatically reduce the required effort. This synergetic business case still under evaluation by both parties.

6. Summary and Conclusion

The work of work package 5 was conceptually divided into two dimensions. On the one hand the dimension of the activity type; we have planned design-related activities (focus groups, design studies) to lay the foundations of the Lea's Box system; we have planned evaluation studies, focussing on the quality of individual system components and designs; finally, we have planned pilots to see how the solutions work in practice on a higher level. On the other hand, there is the dimension of time: In the first project period, the focus was on supporting system design work, in the second period, the focus was on evaluation and piloting work to provide feedback and enable improvements in terms of design and functionality. In the final year, the focus was on supporting the dissemination and exploitation work. In addition, an aim was to present findings that are relevant for the entire Learning Analytics community. These dimensions are reflected in the general evaluation framework we introduced as a response to the second year review (see section 2 of this document). In this framework, the two dimensions are complimented with the dimension of the system components and the dimension of evaluation constructs. In the course of the project we filled the cells of the framework with findings from the project's activities (see next table).

Lea's Box piloting and evaluation results.

	Portal and Appr.	myClass	myActivities	FCA Tool	Mindmap	Learning Spaces	Hasse Diagrams	OLM	Config Tool	Flower App	Interconnection between tools and overall functionality
Functionality and adequacy	D5..3; D5.5;	D5..3; D5.5; Sec. 4.1	D5.5;	Sec. 4.3	D5.3	Sec 5.6	Sec 4.6	D5..3; D5.5; Sec. 5	D5.5;	D5.5; Sec. 3; Sec. 4	D5.5; Sec. 3, Sec. 4.2; Sec. 4.6; Sec. 5
Stability and speed	X	D5.5; Sec. 4.1	X	X	X	X	X	Sec. 5	X	D5.5; Sec. 3; Sec. 4	Sec. 3, Sec. 4.2; Sec. 5
Usability	Sec. 5	D5..3; D5.5; Sec. 4.1	D5.5;	Sec. 4.3	D5.3	Sec 5.6	X	D5..3; D5.5; Sec. 5	D5.5;	D5 D5.5; Sec. 3; Sec. 4	Sec. 3, Sec. 4.2; Sec. 5
Effectiveness, reliability, and validity	D5..3; D5.5;	D5..3; D5.5; Sec. 4.1	X	Sec. 4.3	D5.3	Sec 5.6	D5.3	D5..3; D5.5; Sec. 5	X	D5.5; Sec. 3; Sec.4	Sec. 3, Sec. 4.2, Sec. 4.6; Sec. 5
Logistical and practical obstacles, feasibility and limitations	D5..3; D5.5;	D5..3; D5.5; Sec. 4.1	X	Sec. 4.3	D5.3	D5.3; Sec 5.6	Sec 4.6	D5..3; D5.5; Sec. 5	X	D5.5; Sec. 3; Sec.4	Sec. 3, Sec. 4.2; Sec. 4.6; Sec. 5
Acceptance	D5..3; D5.5;	D5..3; D5.5; Sec. 4.1	X	Sec. 4.3	X	Sec 5.6	X	D5..3; D5.5; Sec. 5	D5.5;	D5.5; Sec. 3; Sec. 4	Sec. 3, Sec. 4.2; Sec. 5

Although the attempt to fill all cells of this matrix is out of the scope of the project, a distinct gap refers to the computational performance. The question whether a system works stable and fast enough is undoubtedly important. To a certain degree we can answer such questions from the usability side; slow and unstable systems are not considered usable by the users. A more systematic approach to load testing was given a low priority.

The work of WP5 can also be summarized in terms of the key performance indicators as of the DoW. The first numbers before the slash indicate the planned studies or participating teachers as of the DoW, the middle numbers indicate the re-planning as of the annex to the second periodic report, and the number after the slash indicate the actual achievements.

Pilot Studies			
	Country	Nr. of Pilot Studies	Nr. of Teachers
Y3	CZ	2/2/4	3 to 18
	TR	3/2/4	up to 100
	AT, DE	-/3/3	1 to 25

Evaluation Studies			
	Country	Nr. of Pilot Studies	Nr. of Teachers
Y3	CZ	5/5/5	3 to 11
	TR	5/6/7	up to 100
	AT, DE	-/3/3	50

These tables show that the project performed well and met its plans. It is difficult, tough, to draw clear borders between the individual activities. Not only that we have a rather complex evaluation framework, schools, teachers, or students participating in pilots, of course, also contributed to the goals of evaluation studies. The distinguishing factor can only be the nature of information that is gained from an activity.

Summarizing the results of the activities in WP5, we can say that they revealed concrete evidence that a Learning Analytics support of teachers lead to better achievement, engagement and stronger agency. However, these benefits come with 2 conditions. Technical perfection and fluency at the first contact with the software are the largest determining factors for the users to adopt analytics frameworks for everyday use. Factors such as an easy URL, easy login, simple use cases, browser support, mobile support affects hugely. The evaluation studies also reveal that peer influence is a great factor in adoption. When students start to talk about the application being “cool” or being “cumbersome,” the idea spreads very easily and becomes a general belief. Part of the project outcome (e.g. the OLM tools) is planned to be released as open source at the end of the project. The open source preparation activities as well as dissemination material creation are planned being aware of these perception factors. Thus, it is critical to offer very simple and very professional (in the sense of usability, stability, configurability, security, etc.) solutions.

The second condition is that the learner, teacher, school, community, society has to be ready to let the analytics influence their learning paths, standards, thinking. As we have seen in some of our countries, the residing habits from the past sometimes obstruct the uptake of learning analytics tools. There is a great need for further work to be done on reshaping assessment cultures in our countries.

Take-away message

The main message of this report is that Learning Analytics is considered a positive and useful thing. However, the possibilities of this technology must be translated into the very concrete needs, mental models, believes, and concerns of teachers. Also, solutions must be tailored to the daily practice of teachers including the fact that classroom scenarios cannot provide clean and rich data sets for analyses. The understanding and the needs of teachers are different from the solutions and ideas and the thinking of researchers, thus the requested solutions are often much simpler than expected or devised by researchers. Thus, big and complex solutions will fail. In all our experience, teachers tend to prefer simple and clean solutions, even if the ‘power’ of such small solutions is limited. The modular approach of Lea’s Box appears being a highly promising approach to bring the concepts and the ideas of Learning Analytics into schools and classrooms. With a ‘flat’ learning curve’, teachers are curious and willing to dig deeper into the genre and use more and more data tools in their daily practice.

REFERENCES

Albert, D., & Held, T. (1999). Component Based Knowledge Spaces in Problem Solving and Inductive Reasoning. In D. Albert & J. Lukas(Eds.), Knowledge Spaces: Theories, Empirical Research, and Applications (pp. 15–40). Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, L. (2013). Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Pearson Education Limited

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.

Bundesministerium für Bildung und Frauen (2012). Die kompetenzorientierte mündliche Reifeprüfung in den Unterrichtsgegenständen - Biologie und Umweltkunde Empfehlende Richtlinien und Beispiele für Themenpool und Prüfungsaufgaben. Available at http://ahs-tirol.tsn.at/www/sites/default/files/LF_m%C3%BCndlich_BIU.pdf (01.12.2016).

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

Felder, R. M., & Silverman, L. K. (1988). Learning styles and teaching styles in engineering education. Engineering Education, 78(7), 674–681.

Felder, R. M., & Spurlin, J. (2005). Applications, Reliability and Validity of the Index of Learning Styles. International Journal of Engineering Education, 21(1), 103-112.

Ginon, B., Johnson, M. D., Turker, A., & Kickmeier-Rust, M. D. (in press). Students' Visualisation Preferences When Using a Persuadable Open Learner Model. To appear in the Journal of Learning Analytics.Heller, J., Steiner, C.M., Hockemeyer, C., & Albert, D. (2006). Competence-based knowledge structures for personalized learning. International Journal on E-Learning, 5, 75-88.

Hughes, G., & Dobbins, C. (2015). The utilization of data analysis techniques in predicting student performance in massive open online courses (MOOCs). Research and Practice in Technology Enhanced Learning, 2015 10:10.

Kim, J. H., Park, Y., Song, J., Jo, I.-H. (2014). Predicting Students' Learning Performance by Using Online Behavior Patterns in Blended Learning Environments: Comparison of Two Cases on Linear and Non-linear Model. In proceedings of the International Conference on Educational Data Mining (EDM 2014), London, UK.

Papamitsiou, Z., & Economides, A. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. Educational Technology & Society, 17 (4), 49–64.

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.

Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A Review on Predicting Student's Performance Using Data Mining Techniques. Procedia Computer Science, 72, 414-422.

APPENDIX

LEARNING STYLES INVENTORY (GERMAN VERSION)

Liebe Teilnehmerin, lieber Teilnehmer!

Die folgenden Aussagen beziehen sich auf das Lernen, bzw. welche Lernsituationen oder Lernressourcen man bevorzugt. Alle Aussagen beginnen gleich; spalten sich dann allerdings in zwei alternative Enden auf. Wähle bei jeder Aussage jenes Satzende aus welches für dich am eher zutrifft - es gibt keine richtigen oder falschen Antworten. Solltest du ein Fremdwort nicht kennen oder verstehen, dann frage bitte einfach deinen Lehrer.

Deine „ID“:____________________

1. Ich verstehe vieles besser nachdem ich es…

O ausprobiert habe.

O durchgedacht habe.

2. Ich würde eher als...

O realistisch eingeschätzt werden.

O innovativ eingeschätzt werden.

3. Wenn ich darüber nachdenke was ich gestern gemacht habe, würde mir eher…

O ein Bild einfallen.

O etwas Geschriebenes einfallen.

4. Ich tendiere dazu…

O die Details eines Themas zu verstehen aber nicht das "große Ganze".

O das "große Ganze" zu verstehen aber nicht dessen Details.

5. Wenn ich etwas Neues lerne hilft es mir…

O darüber zu reden.

O darüber nachzudenken.

6. Wenn ich ein Lehrer / eine Lehrerin wäre würde ich einen Gegenstand lehren…

O in dem es um Fakten und Beispiele "aus dem echten Leben" geht.

O in dem es um Ideen und Theorien geht.

7. Neue Informationen bekomme ich bevorzugt…

O mittels Bildern, Diagrammen, Graphen oder Karten.

O in geschriebener oder mündlicher Form.

8. Sobald ich…

O die einzelnen Teile verstanden habe, verstehe ich das Ganze.

O das Ganze verstanden habe, verstehe ich wie sich die einzelnen Teile zusammenfügen.

9. Wenn in einer Lerngruppe an einem schwierigen Teil gearbeitet wird…

O trete ich in den Vordergrund und trage aktiv bei.

O trete ich in den Hintergrund und höre zu.

10. Ich finde es einfacher…

O Fakten zu lernen

O Konzepte zu lernen.

11. In einem Buch mit vielen Abbildungen und Diagrammen…

O schaue ich mir die Abbildungen und Diagramme sehr genau an.

O fokussiere ich auf den geschriebenen Text.

12. Wenn ich mathematische Probleme löse…

O arbeite ich mich Schritt für Schritt an die Lösung heran.

O weiß ich oft schon wie die Lösung aussehen wird und muss dann herausfinden welche Schritte ich machen muss um zu dieser Lösung zu kommen.

13. Wenn ich Sachbücher lese, bevorzuge ich es…

O wenn sie Fakten oder Anleitungen enthalten wie ich etwas tun kann.

O wenn sie mir neue Ideen vermitteln über die man nachdenken muss.

14. Ich mag Lehrer / Lehrerinnen…

O die mehr Zeit darauf verwenden um Diagramme und Abbildungen auf die Tafel zu zeichnen.

O die mehr Zeit darauf verwenden um Lerninhalte mündlich zu erklären.

15. Wenn ich eine abstrakte Geschichte oder Erzählung analysiere…

O denke ich an die einzelnen Wendungen und Ereignisse um die grundlegenden Themen und Motive zu verstehen.

O kenne ich die grundlegenden Themen und Motive wenn ich fertiggelesen habe und muss dann die Wendungen und Ereignisse finden um meine Annahmen zu untermauern.

16. Wenn ich an einer Hausübung arbeite versuche ich…

O direkt die Lösung herauszufinden.

O das Problem zuerst vollends zu verstehen.

17. Ich bevorzuge…

O Gewissheiten, also Themen die allgemein als „wahr“ oder „falsch“ angesehen werden.

O Themen die noch umstritten sind, als noch nicht allgemein als „wahr“ oder „falsch“ angesehen werden.

18. Ich kann mir am besten merken…

O was ich sehe.

O was ich höre.

19. Für mich ist es wichtiger wenn ein Lehrer / eine Lehrerin…

O das Lernmaterial chronologisch, d.h. Schritt für Schritt vorgibt.

O einen Überblick gibt und das Lernmaterial mit anderen Gegenständen oder Themen in Beziehung setzt.

20. Ich bevorzuge…

O in einer Gruppe zu lernen.

O alleine zu lernen.

21. Ich würde eher als jemand eingeschätzt werden der / die…

O sich gewissenhaft um die Details einer Aufgabe kümmert.

O sich kreativ damit auseinandersetzt auf welche Art die Aufgabe gelöst werden kann.

22. Wenn ich einen unbekannten Ort aufsuchen muss…

O bevorzuge ich eine Karte.

O bevorzuge ich eine Wegbeschreibung.

23. Ich lerne…

O zeitlich mit relativ konstanter Geschwindigkeit. Wenn ich hart und lange genug gelernt habe verstehe ich es.

O mit vielen Pausen und Unterbrechungen. Ich bin oft komplett verwirrt aber plötzlich geht mir ein Licht auf.

24. Ich würde…

O damit beginnen die Dinge einfach auszuprobieren.

O zuerst darüber nachdenken wie ich es angehen kann.

25. Wenn ich in meiner Freizeit lese bevorzuge ich Schriftsteller / Schriftstellerinnen…

O die sich klar und schnörkellos ausdrücken.

O die sich in einer kreativen, interessanten Sprache ausdrücken.

26. Wenn ich im Unterricht ein Diagramm oder eine Skizze sehe erinnere ich mich später eher…

O an genau diese Abbildung.

O was der Lehrer / die Lehrerin dabei erklärt hat.

27. Wenn ich mit einer Vielzahl an Informationen konfrontiert werde…

O versuche ich mich zuerst auf die Details zu konzentrieren um das "große Ganze" zu verstehen.

O versuche ich zuerst das "große Ganze" zu verstehen und gehe dann dazu über die Details zu lernen.

28. Ich kann mich besser an etwas erinnern…

O das ich getan oder ausprobiert habe.

O über das ich länger und ausführlich nachgedacht habe.

29. Wenn ich eine komplexe Aufgabe zu lösen habe bevorzuge ich es…

O bereits bekannte Lösungswege anzuwenden.

O neue Lösungswege herauszufinden.

30. Wenn mir jemand neue Informationen präsentiert…

O bevorzuge ich Ergebnisse in Form von Diagrammen oder Abbildungen.

O bevorzuge ich eine schriftliche Zusammenfassung der Ergebnisse.

31. Wenn ich einen Aufsatz schreiben muss...

O schreibe ich in der vorgesehenen Reihenfolge, also zuerst Beginn, dann Mittelteil, dann Ende.

O schreibe ich die einzelne Teile in einer beliebigen Reihenfolge und ordne sie dann zum Schluss.

32. Bei einer größeren Gruppenarbeit / einem größeren Gruppenprojekt bevorzuge ich es wenn….

O die Gruppe als Ganzes ein "Brainstorming" durchführt bei dem die Ideen Aller gesammelt werden.

O die Gruppenmitglieder für sich alleine ein "Brainstorming" machen um dann in der Gruppe die Ideen zu sammeln und zu vergleichen.

33. Ich denke es ist größeres Lob wenn man jemanden…

O rational nennt.

O einfallsreich nennt.

34. Wenn ich auf einer Party neue Leute treffe erinnere ich mich eher daran…

O wie sie ausgesehen haben.

O was sie über sich erzählt haben.

35. Wenn ich mit dem Lernen eines neuen Themengebietes beginne dann bevorzuge ich es…

O mich nur auf dieses Themengebiet zu konzentrieren und darüber so viel wie möglich zu lernen.

O mich auch über andere verwandte Themengebiete zu informieren um Verbindungen zwischen diesen und dem eigentlichen Themengebiet herzustellen.

36. Ich glaube von Anderen eher als…

O extrovertiert eingeschätzt zu werden.

O introvertiert eingeschätzt zu werden.

37. Ich bevorzuge Unterrichtsfächer in denen es…

O um etwas Konkretes geht (Fakten, Daten).

O um etwas Abstraktes geht (Konzepte, Theorien).

38. Zur Unterhaltung würde ich eher…

O fernsehen.

O ein Buch lesen.

39. Manche Lehrer / Lehrerinnen beginnen ihren Unterricht mit einem Überblick über das zu behandelnde Thema. Solch einen Überblick...

O finde ich einigermaßen hilfreich.

O finde ich sehr hilfreich.

40. Der Idee, Hausaufgaben in der Gruppe zu machen, wobei die Gruppe als Ganzes benotet wird…

O kann ich etwas abgewinnen.

O kann ich nichts abgewinnen.

41. Wenn ich langwierige Berechnungen durchführe…

O tendiere ich dazu alle Rechenschritte zu wiederholen um die Berechnung zu überprüfen.

O finde ich es mühsam meine Ergebnisse zu überprüfen und muss mich zwingen es zu tun.

42. Ich kann mir neue Orte / Plätze…

O sehr leicht und gut einprägen.

O nur schwer und ohne viele Details merken.

43. Wenn im Zuge einer Gruppenarbeit / eines Gruppenprojekts ein Problem gelöst werden soll, dann…

O denke ich eher an die einzelnen Schritte im Lösungsprozess.

O denke ich eher an die möglichen Konsequenzen der Lösung in unterschiedlichen Anwendungsgebieten.

INTRINSIC MOTIVATION INVENTORY (ORIGINAL)

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

INTRINSIC MOTIVATION INVENTORY (GERMAN VERSION)

Liebe Teilnehmerin, lieber Teilnehmer!

Deine „ID“:____________________ Datum:_________________

Im Projekt der vergangenen Wochen hast du dich in einem mit den Themen „Ökologie und Ökosysteme“ auseinandergesetzt. Nun beginnt ein neues Projekt zum Thema „Die Sinne“.

Dabei wirst du insgesamt dreimal gebeten werden einen Fragebogen – wie du ihn unten siehst – auszufüllen. Jetzt, also in der ersten Runde, geht es um deine Erwartungen wie das Projekt sein wird – denn noch hat es ja nicht begonnen. Es gibt keine richtigen und falschen Antworten!

In der zweiten und dritten Runde wird es dann um deine tatsächlichen Erfahrungen mit dem Projekt gehen, denn dann wird es schon mehr oder weniger fortgeschritten sein.

Bitte gib für jede Aussage an inwieweit sie auf dich zutrifft. Sollte eine Aussage überhaupt nicht auf dich zutreffen, dann kreuze das Kästchen mit der „1“ an. Solltest du einer Aussage absolut zustimmen, dann kreuze das Kästchen mit der „7“ an.

	Überhaupt nicht						Absolut
1. Ich glaube das wird eine langweilige Aufgabe werden.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
2. Ich glaube dass die Bearbeitung dieser Aufgabe nützlich sein wird.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
3. Ich habe vor mich sehr anzustrengen um die Aufgabe gut zu erfüllen.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
4. Die Bearbeitung der Aufgabe wird mir Spaß machen.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
5. Ich glaube dass es mir Vorteile bringen wird an dieser Aufgabe zu arbeiten.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

6. Es ist wichtig für mich gut abzuschneiden.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
7. Ich glaube die Aufgabe wird interessant sein.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
8. Ich glaube dass es mir einen Nutzen bringen wird an dieser Aufgabe zu arbeiten.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
9. Ich habe vor mich sehr einzusetzen.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

LEARNING LOG-BOOK (GERMAN VERSION)

Liebe Teilnehmerin, lieber Teilnehmer!

Auf den folgenden Blättern findest du eine Tabelle. Diese Tabelle besteht aus 4 Spalten mit den Überschriften: „Datum“, „Lernressource“, „Stichwortartige Zusammenfassung“ und „Bewertung“.

Diese Tabelle dient sozusagen als dein „Lerntagebuch“ in den kommenden Wochen. Bitte trage in jeder neuen „Theorie-stunde“ das Datum in die erste Spalte ein.

Trage bitte in der zweiten Spalte gut leserlich den Titel der „Lernressource“ ein die du gerade eben gelesen, angeschaut, gelernt oder erkundet hast.

Trage bitte in der Spalte „Stichwortartige Zusammenfassung“ stichwortartig ein paar Begriffe auf, die *deiner Meinung nach* die Lernressource am besten zusammenfassen.

In der Spalte „Bewertung“ kannst du dieser Lernressource „Punkte vergeben“, und zwar von 0 bis maximal 5 Punkten.

Bewerte bitte inwieweit dir die Lernressource gerade eben geholfen hat das Thema des Projekts (also „Die Sinne“) noch besser zu lernen und zu verstehen.

Vergib 0 Punkte wenn dir die Lernressource gerade eben überhaupt nicht geholfen hat und bis zu 5 Punkte wenn die Lernressource in dieser Hinsicht perfekt war.

Hier gibt es keine richtigen oder falschen Antworten!

Es geht ja nicht darum zu bewerten wie gut oder schlecht die Lernressource an sich ist, sondern wie hilfreich sie für dich gerade eben war. Um diesen Unterschied an einem Beispiel zu erklären: Eine tolle, stark vergrößerte Abbildung von den Facettenaugen einer Fliege kann

an sich eine ausgezeichnete Lernressource sein, aber wenn ich noch kaum etwas über Insekten oder Fliegen im Speziellen weiß wird sie mir* in diesem Moment nicht so gut dabei helfen das Thema „Insekten“ oder Fliegen zu verstehen und zu lernen.

*(mir persönlich…denn Lernende sind unterschiedlich und haben unterschiedliche Vorlieben wie sie ein Thema erlernen).

Deine „ID“:____________________

Datum (Date)	**Lernressource** (Learning Ressource)	**Stichwortartige Zusammenfassung** (Summary in note-form)	**Bewertung von 0 bis 5 Punkte** (Evaluation from 0 to 5 points)

Learning Resource Evaluation Form for the Teacher

active	**reflective**
Innovativ, offen für neue Wege, unkonventionell „aktives Lernmaterial“ / ausprobieren Kommunikation mit anderen / kann oder soll mit anderen gelernt werden	konservativ, nicht offen für neue Wege, konventionell „passives Lernmaterial“ / konsumieren Soll eher alleine gelernt werden
sensing	**intuitive**
Fakten Konkretes Material Praktisch – „Real world“	Theorien, Möglichkeiten, „Strittiges“ Abstraktes Material „Bildung“ im Humboldtschen Sinne
visual	**verbal**
Bilder Diagramme Mindmaps	Texte Mündliche Erklärungen
sequential	**global**
Lernen in kleinen Lernsequenzen Schritt bei Schritt Kleine „Häppchen“	„Holistisches“ Lernen („den Wald statt einzelne Bäume sehen“) Überblick verschaffen „Unlogische“ Abfolge von Lernressorcen („mal dies, mal das“) Verbindungen zwischen einzelnen Blöcken werden betont

Lernressource:

active	O (0)	O (1)	O (2)	O (3)	O (4)	reflective
sensing	O (0)	O (1)	O (2)	O (3)	O (4)	intuitive
visual	O (0)	O (1)	O (2)	O (3)	O (4)	verbal
sequential	O (0)	O (1)	O (2)	O (3)	O (4)	global

Lernressource:

active	O (0)	O (1)	O (2)	O (3)	O (4)	reflective
sensing	O (0)	O (1)	O (2)	O (3)	O (4)	intuitive
visual	O (0)	O (1)	O (2)	O (3)	O (4)	verbal
sequential	O (0)	O (1)	O (2)	O (3)	O (4)	global

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

TECHNOLOGY ACCEPTANCE MODEL 3 (TAM3)

TAM3 evaluations are done in four stages:

Stage 1 – Run a demographic segmentation survey: Of the pilot participants' data, cuts for different populations will be produced (attending residential vs. commuter institutions, attending elite vs. less selective schools, STEM vs. humanities and social science majors, etc.) The segmentation profiling different populations at different kinds of institutions will provide useful data, information, and insight into the aggregate experience of participants as well as the variations in the experiences of different profiles.

Stage 2 – Allow enough time with the system: The information system whose acceptance is to be modelled has to be used long enough by the users to have the users shape their ratings to statements such as "Using the system would improve my job performance." Although the impact of eLearning systems on performance takes many years, basic indicators for positive affect surface in typically 4 weeks. This is because involving a technology component in a social process would naturally transform that process which happens in time. TAM3 reveals which aspects of the system are critical in shaping users' intentions to sustain the actual use of the system for so long.

Stage 3 – TAM3 Survey: A number of TAM3 models applied to users' intentions towards eLearning technologies sought an answer to the following questions:

1. Does a relationship exist between subjective norm and perceived usefulness?
2. Does a relationship exist between subjective norm and intention to use technology?
3. Does a relationship exist between perceived ease of use and perceived usefulness?
4. Does a relationship exist between perceived usefulness and intention to use technology?
5. Does a relationship exist between perceived ease of use and intention to use technology?

The outcomes prove that user's positive feeling towards the ease of use of technology is associated strongly with the sustained use of the technology. Moreover, perceived ease of use (PEU) has a significant influence on perceived usefulness (PU). This is one hypothesis that will be checked in Lea's Box pilots. For the Lea's Box, a standard TAM survey will be used where a set of psychometric statements are asked to be rated on a 7 level Likert scale, plus an option for "No Opinion." These statements are statistically proven by Davis (1989) to have enough coverage on the variables to be measured. The standard TAM survey is as follows:

PERCEIVED USEFULNESS		1	2	3	4	5	6	7		NA
1. Using the system in my job would enable me to accomplish tasks more quickly	unlikely	○	○	○	○	○	○	○	likely	○
2. Using the system would improve my job performance	unlikely	○	○	○	○	○	○	○	likely	○
3. Using the system in my job would increase my productivity	unlikely	○	○	○	○	○	○	○	likely	○
4. Using the system would enhance my effectiveness on the job	unlikely	○	○	○	○	○	○	○	likely	○
5. Using the system would make it easier to do my job	1unlikely	○	○	○	○	○	○	○	likely	○
6. I would find the system useful in my job	unlikely	○	○	○	○	○	○	○	likely	○
PERCEIVED EASE OF USE		**1**	**2**	**3**	**4**	**5**	**6**	**7**		**NA**
7. Learning to operate the system would be easy for me	unlikely	○	○	○	○	○	○	○	likely	○
8. I would find it easy to get the system to do what I want it to do	unlikely	○	○	○	○	○	○	○	likely	○
9. My interaction with the system would be clear and understandable	unlikely	○	○	○	○	○	○	○	likely	○

10. I would find the system to be flexible to interact with	unlikely ◌	◌	◌	◌	◌	◌	◌	likely ◌
11. It would be easy for me to become skillful at using the system	unlikely ◌	◌	◌	◌	◌	◌	◌	likely ◌
12. I would find the system easy to use	unlikely ◌	◌	◌	◌	◌	◌	◌	likely ◌

Stage 4 – Factor Analysis: The data collected is placed in a correlation matrix and values of correlations between variables and those of the partial correlations are compared by Kaiser-Meyer-Olkin (KMO) measure to check if the samples can be considered "meritorious" to carry out factor analysis. Subsequently, carry out the factor analysis of the 5 variables that load "perceived usefulness" and the 4 variables that load "perceived ease of use" for the evaluated system.