iTalk2Learn
31-10-2013

Deliverable 7.2

Report on Project Evaluation Plan

31 October 2013

Project acronym: iTalk2Learn

Project full title: Talk, Tutor, Explore, Learn: Intelligent Tutoring and Exploration for Robust Learning

Work Package: **7**

Document title: Report on Project evaluation Plan

Version: **1.2**

Official delivery date: **31.10.2013**

Actual publication date:

Type of document: Report

Nature: Public

Authors: Carlotta Schatten with the help of the partners

Approved by: _____

Version	Date	Sections Affected
1.0	07.10.2013	Initial version
1.1	21/10/2013	Review comments processed BBK, RUB and Whizz
1.2	28/10/2013	Final review and submission

Executive Summary

In this deliverable we will discuss metrics of success for the project. We will discuss the internal and external metrics, together with the impact that the project will have on the market. Direct connections are made to WP5 and WP6 deliverables that contain more detailed descriptions of the success pursuing activities that have been completed and planned by the consortium.

Table of Contents

1. Introduction

In order to pursue and evaluate the objectives of the project it is necessary to define an evaluation plan, a dissemination plan, and an exploitation plan. The evaluation plan as part of WP5 is described in D5.1. The plans for dissemination and exploitation as part of WP6 are described in D6.3.1 and D6.4.1. In this deliverable we will define qualitative and quantitative metrics of success in respect to the mentioned objectives in the DOW.

We measure the results of the project, with respect to 4 objectives
1) Provide an open-source platform for intelligent support systems integrating structured practice and exploratory, conceptually-oriented learning.
2) Provide state-of-the-art and highly innovative reference implementations of plugins for the platform that could be used in a wide range of application domains.
3) Promote our understanding of the role of the different modalities of speech and direct manipulation and multiple and alternative representations in learning elementary mathematics through digital technologies.
4) A summative evaluation of activities and support features generated by our intelligent learning support platform in two field trials in two different settings

Moreover, we distinguish between internal and external success measures. Internal success measures are defined in order to institute an internal mutual control between the partners and to get an overview of the situation of each WP. The scope of external metrics of success, instead, is to get a step forward in the research areas involved. They are called external because the success is measured by persons who are not member of the consortium, whereas internal metrics involve unofficial advices and measurements of the iTalk2Learn partners. What we mean in detail will be explained in the following sections.

2. Internal Metrics of Success

In this section we discuss how the consortium, through its activities, evaluates the participating members.

In the monthly teleconferences and general project meetings, the partners are involved in discussions about the work done by the partners. It is of primary importance for each of the partners to have the approval from the other members of the consortium, because they count as experts in their area and because the work for WP is strongly interdisciplinary.

We evaluate the contribution of each partner through the quality and quantity of work done and

punctuality. In order to keep track of this, each presentation delivered is collected in the project file sharing platform Owncloud.

A mutual control is also exercised in the deliverable production, whereby for each deliverable, two internal reviewers are tasked to evaluate the work and propose any necessary changes. For each deliverable, we select one partner with an expertise related to the work described in the deliverable and one partner whose expertise is only tangent to the described work. In this way we guarantee that the deliverables have a high standard concerning both the in-depth content as well as the understandability for other researchers.

We have set up a dissemination and exploitation plan as a live document in Google Docs. This is used to measure the progress in technical dissemination and exploitation. All partners are asked to contribute to this as an ongoing requirement. This will aid us in evaluating the types of dissemination and exploitation activities that are being planned or have taken place, as well as the success of their implementation. This is a key way in which we will measure the success for WP6 and the commitments that we have made within the deliverables in this work package.

Internal success measures of the cognitive and pedagogical models of the learning process in fractions (called intervention model, see D5.1) will be gathered from existing literature sources and curricula from England and Germany. The intervention model will inform the summative evaluation studies as described in D5.1. Internal evaluation related to the success of cognitive models will come from peer reviews mostly by colleagues who have been involved in research of this domain previously. Peer reviews of the pedagogical models will be facilitated by project partners Whizz and their connections with schools and educational consultants. Further reviews of both the cognitive and pedagogical models will come from 'critical friends' of the project, such as teachers or other researchers closely attached to the project, a professor of mathematics education from another institution, who are acting as advisors throughout the design stages. A similar role is done officially by the advisory board. As promised in the DOW we will meet them twice during the lifetime of the project. Those official meetings will be planned for the next two years. In the first year this was done in an informal way.

Internal evaluation of the success of the design of the tasks, sequences and interventions will occur through on-going small-scale trials involving project researchers and students. The results from these formative evaluations will provide continuous feedback to enhance further design iterations throughout the project. About that more details can be found in D5.1.

A final internal evaluation will be carried out towards the end of the project; the success of iTalk2Learn will be internally evaluated according to the extent to which the data collected can be used. The collected data is of primary importance for machine learning based technologies as the ones of Sail and all the tasks of WP2 that involves the University of Hildesheim and Birbeck Collage. This is due to the fact that the machine learning algorithms learn from them and the more data are present the better they can work. Since the tasks suggested are highly innovative, no state of the art data sets are available for some of the tasks promised, e.g. speech recognition of German and

English (UK) children, performance prediction in exploratory tasks, switching between exploratory and structured activities. Consequently not only the exploitation of the work done in the project depends from the data collected, but also the success of the tasks.

3. External Metrics of success

In this Section we discuss the metrics of success that involves judgment of expert who are not directly involved in the project.

3.1 General metrics

For each University partner the number of publications is generally a good success measure.

As written in the dissemination deliverable, we will try to publish in the major international conferences, such as AERA, AIED, CSEDU, EARLI, EC-TEL, EDM, ICLS, IEEE ICALT, ITS, KDD, RecSys, UMAP, ICASSP, INTERSPEECH and LREC. All of these conferences have a strict peer review procedure. We will also submit manuscripts to international research journals, such as Computers & Education, IEEE TLT, IS, JCAL, JLS, JUCS, L&I, UMUAI.

For SME partners, instead, the metric of success is defined in how they will exploit the work done in the project to expand into new markets in the EU (and possibly outside of the EU). The SME partners will try to strengthen their position in the global market and bring the same state-of-the-art learning technology to all members of the EU, helping to shorten the skills and economic gap between the member countries.

However, the metrics for success in terms of exploitation for each individual partner will differ depending upon the goals for their organisation and the perspectives from which they are approaching the project. See D6.4.1 for more details.

3.2 Field trials metrics

Towards the end of the project it will be possible to measure the level of some of the stakeholders' acceptance of our technology through the following.

1. Acceptance of the learning platform by the students who are using and trialling the technology for the project: their opinion is a crucial element; we are committed to taking it into account, and this is a strong external measurement of the project's success.
2. Evaluation by teachers (through questionnaires or interviews) analysing how the learning platform addresses the cognitive and pedagogical development. We will investigate which aspects teachers, as experts of elementary mathematics teaching, would champion.

Furthermore, success will be measured by evaluating the general uptake of the learning platform by teachers and students but this will be visible only after the project. All the two mentioned feedback

forms and surveys will pose questions on the two aforementioned project objectives in order to evaluate their success separately. This feedback would then be analysed in order to reach a conclusion as to the success of the platform. However, any conclusions as to the acceptance of the technology into the wider market, i.e. outside the schools that are involved in the trials, will be purely hypothetical. This will enable us to gauge the success of the iTalk2Learn platform and compare this to the success of the Whizz tutoring system in a limited environment and with limited testers, but it will not indicate how likely the market is to adopt the platform thereafter. How this is done is explained in detail in D6.4.1.

Therefore, another key evaluation metric will be the extent to which the iTalk2Learn message has been spread throughout the market. Please see D6.3.1 for details of these dissemination efforts and their importance. In order to promote the acceptance of our technology by the educational community, the community must be informed about the benefits of such systems. This is something that Whizz currently does for online tutoring by taking the system to classes and explaining its benefits. Please see D6.4.1 for more information on the importance of educating the market in order to improve the project's potential for success.

4. Impact on the Market

The impact on the market can be interpreted as a general success metric for iTalk2Learn platform. Although we will not be able to fully evaluate it during the project, we expect the technology to have an impact upon the market in a number of ways. This will be initially outlined in the analysis of the market situation, potential clients etc. in D6.4.1, and will be expanded upon in D6.4.2 and D6.4.3 in subsequent years of the project. However, a provisional and rough outline of the expectation of the impact upon the market is highlighted below.

In order to understand the potential impact of the technology on the market, it is important to first understand the background of the market perspective. The education market, and therefore the potential clients of the iTalk2Learn platform, may not be familiar with the concept of online tuition and intelligent tuition. The more the message is spread and the market is educated, the more likely it is that the technology will be more widely accepted in the market. The potential impact of this technology is growing. Online tuition is predicted to become more and more important in the students' life.

The expected impact would be for more children to have access to personalised online assessments and tuition. The desired outcome is that each of these students should be more likely to reach their potential in the areas in which they are being tutored.

This would essentially replicate the benefits of a human tutor. However, the impact upon the market is much greater than that, since the online solution is much more scalable. The subset of

students that would have access to private tuition is much higher when online tuition is introduced – private tuition is no longer limited to those that can afford expensive tuition fees and those that are not embarrassed to admit that they need extra support. The iTalk2Learn platform could have the potential to rapidly increase the percentage of students that have access to personalised tuition.

In addition to private tuition, the iTalk2Learn platform has the potential to be utilised within the school environment. Schools would be able to provide a healthy amount of individual tuition and attention to all of their students. Many schools currently struggle to provide the individual attention that their students need – often with large class sizes and a wide range of ability within each class – and so the personalised online tuition that the iTalk2Learn platform aims to provide will also broaden the market within schools.

The findings from the iTalk2Learn platform could be applied to online tutoring in subjects other than mathematics. If this were to be extended to other subjects, this technology has the potential to provide every student with tuition in core subjects. Since effective online tuition has already been shown to have a profound impact upon achievement levels, such wide availability of tuition thereby has the potential to profoundly alter and advance the education of students.

It is crucial to remember that the extent of the impact upon the market is dependent upon the way in which the platform is implemented. It must be combined with other services, e.g. overview on the achieved results, reward activities etc., in an effective manner in order to reach a wider market and prove to have a successful impact. As a consequence this fact can be interpreted as a success metric for the iTalk2Learn platform.

5. Conclusion

The project is designed such that measurement of progress against objectives will be done at various times during the project execution. The already started submission of the deliverables will continue during the project and, consequently, facilitate project evaluation and review. A dissemination and exploitation plan is now available and will help the mutual control between partners. It will be kept as a live document that will measure the progress in technical dissemination and exploitation. The management, by means of monthly teleconferences and general meetings, will use the internal metrics to increase the quality of the work done by partners.