Learning FootPrint: A Toolkit for the Visual Analytics of Learning Traces using Heatmaps

Motivation & Introduction

Each student during a lecture takes from the teacher his experience. Then, during the exercise solving he applies this experience and learns new materials. Often in the processes of solving involved other students. In this case students begin to share experiences with each other. Also a lot of the time the student is training independently or discussing material with other students [2]. The speed of such training depends on the style of information retrieval. For example, if a student reads book he can read it from the end or only the first few words in each paragraph. Such processes of finding specific information and separation main things from the secondary is very valuable.

The key ingredient in self-organized learning is the shift of responsibility for the learning process from the teacher to the student. This involve students acquiring an understanding of their learning, being motivated to learn and collaborating with teachers to structure their learning environment. They found a consensus in the literature that independent learning did not merely involve students working alone; teachers have a key part to play in enabling and supporting independent learning though, for example, structuring group work [1]. Here are the advantages of self-organized learning:

- improved academic performance;
- increased motivation and confidence;
- greater student awareness of their limitations and their ability to manage them;
- enabling teachers to provide differentiated tasks for students;
- and fostering social inclusion by countering alienation.

With visual learning analytics we are able to improve self-organized learning. Imagine that the student knows what part of the book has been read the most times. Based on this information, student can search for the information much faster. Such information may also be useful for the teacher. What part of the book has been read the most times? What part of the book, students did not read at all? After gathering all this information about educational materials, video lecture, wiki pages, teacher can make a lot of useful conclusions.

With self-organized learning people handle a huge amount of information. Brain is quite limited in the exact memorization of information, so while we are reading the book we can’t remember a sequence of words or repeat exactly the entire paragraph. After reading each paragraph, we are trying to create conclusions and form a compressed sense in a few key words. That is why if you read not very carefully you must re-read the entire paragraph

from the beginning to get the meaning and leave a main sense in the memory. After reading a large section, we link the information (clustering), and keep in mind. That is why it is so easy to read book about some adventures, such book is the sequence of events that's very easy to relate to each other and remember.

Reading the scientific literature in the context of solving a particular problem is significantly different from the usual reading books. When we solve the task we are faced with a problem, then we try to find an answer in the slides of the lecture, recorded on video lectures or educational materials. This usually occurs in the following sequence:

1. Lack of knowledge
2. Searching in the materials
3. Solution

In this sequence the process of learning takes place at the second point, when a student is faced with a new challenge for himself and looking for the answer in the literature, Internet or video lessons. In order to solve the same problems, students of one group will look for the same information sources. Each student will search answers to questions using their experience. Such information retrieval can be called “trace of learning”. This trace is very useful for teacher and other students. For example, teachers can understand how useful materials for students, or other students can accelerate the search for information in the textbook.

Collection and analysis of the learning process is not an easy task because it is very difficult to track the activity of the student in the internet and even more difficult to do this outside of the Internet - in real life. But today all student learning activity is concentrated in one place - e-Learning systems, so now it becomes possible. After receiving the assignment the student uses a system that gives him access to the virtual classroom with recorded on video lectures, assignments, slides, pdf books, articles, etc. In a centralized e-Learning system analysis of the student learning behavior become real.

Aim of the Thesis & Research Questions

The main goal of this work is to analyze behavior of students while they learning and make it available to other students and teachers. Explore how they read books, which pages were read most times and which opened very rarely. Also in focus are other ways to obtain information such as viewing the video lectures. What moments lectures the most popular, and which are often skipped.

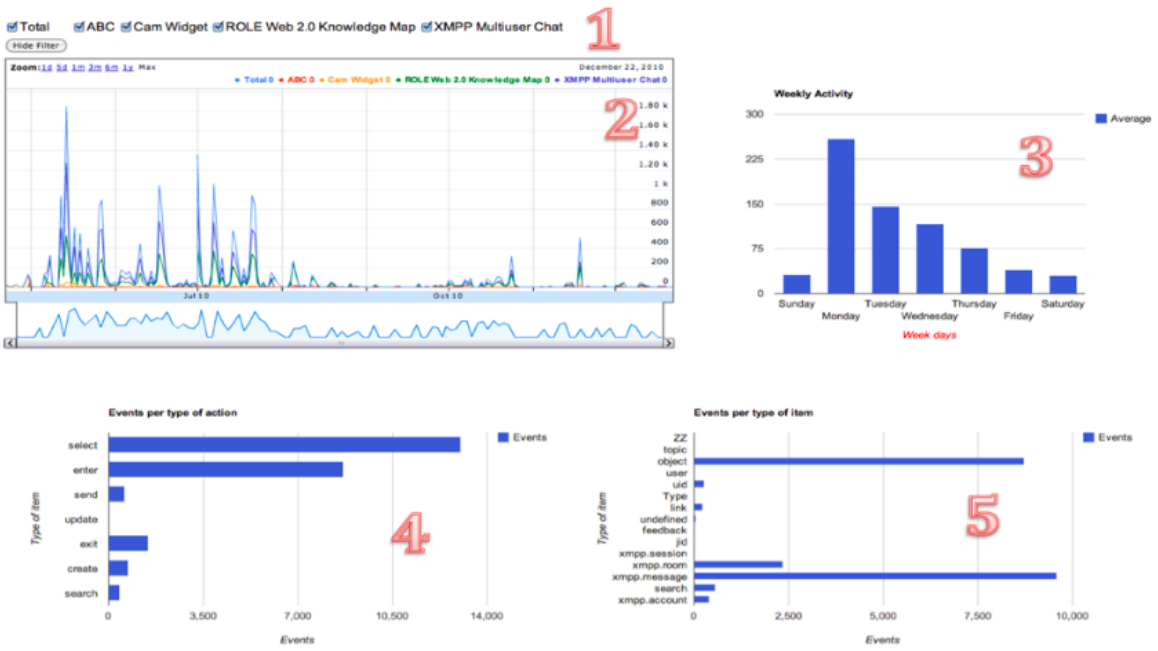
Main research questions:

1. **How to harness learners’ traces to deal with the information overload problem in learning environmrnts?**

2. **How visual learning analytics with heatmaps can support self-organized learning?**

Related work

Learning dashboard - system which shows student activities. For learners and teachers alike, it can be extremely useful to have a visual overview of their activities and how they relate to those of their peers or other actors in the learning experience. In fact, such visualizations can also be quite useful for other stakeholders, like for instance system administrators. [4] Here is example of such system:



In this example dashboard shows how often students use system and some information about actions.

Advantages:

+ Data visualization
+ Variety of the data

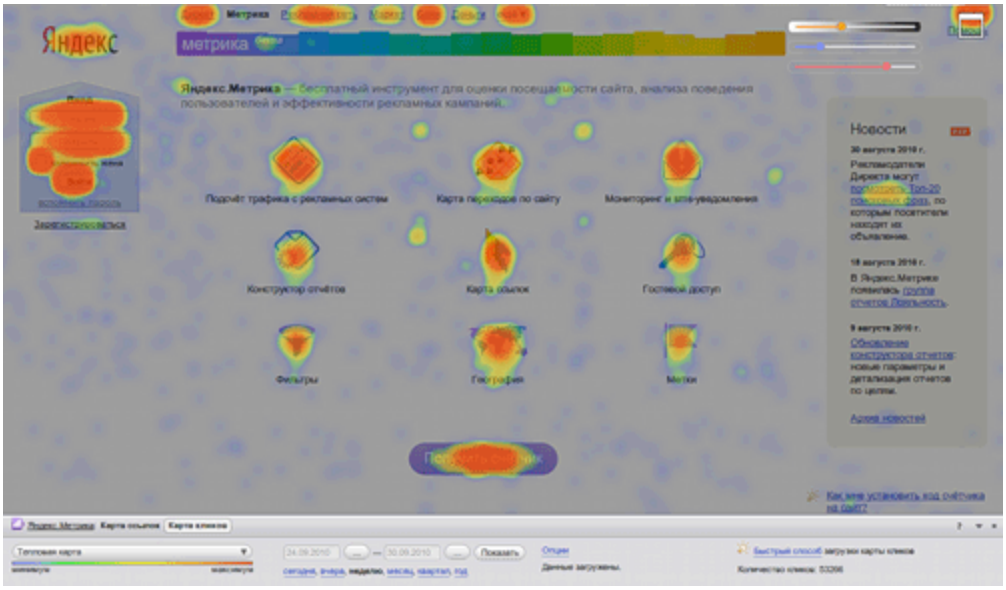
Disadvantages:

- Difficult to use during the learning process
- Difficult to integrate with other systems

There are many examples of systems that make the analysis of user behavior on the Internet. One of the biggest projects is Google Analytics. It is a system that keeps track of

users by collecting statistical information. Location, gender, approximate age and most importantly - behavior users on the site. For example, you can learn how often users click on the button to buy and what is even more important how often they do not press and for what reason.

Another well-known project is Yandex.Metrics. It monitors all movements of the mouse on the website, and based on this information generate the heat map of the web site:



Such information is very valuable for people who are engaged in business. Also it helps interface designers. For example, you made a new payment button and want to test how users react to it, it's all very simple to do with Yandex.Metrics.

Advantages:

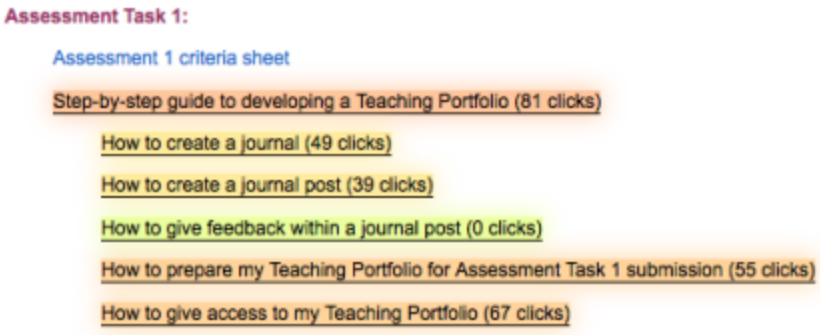
+ Show the behavior of users
+ Tools for analysis
+ Easy to integrate

Disadvantages:

- Built for Business
- Not visible for content users

All of these systems are aimed at business and display all this information is only for owners of web sites. Such a system can’t help to share the most interesting places of the page to the all user of the website.

The Moodle Activity Viewer (MAV) – Heatmaps of Student Activity shows how often the students click on the link. If the link popular then the background color will be red, unpopular links have green background. Such information is useful but it does not say anything about the content. We can find out what content is the most interesting and important but we can not say which part of it. Here's a screenshot of the system:



Advantages:

+ Shows the importance of links
+ Heatmaps

Disadvantages:

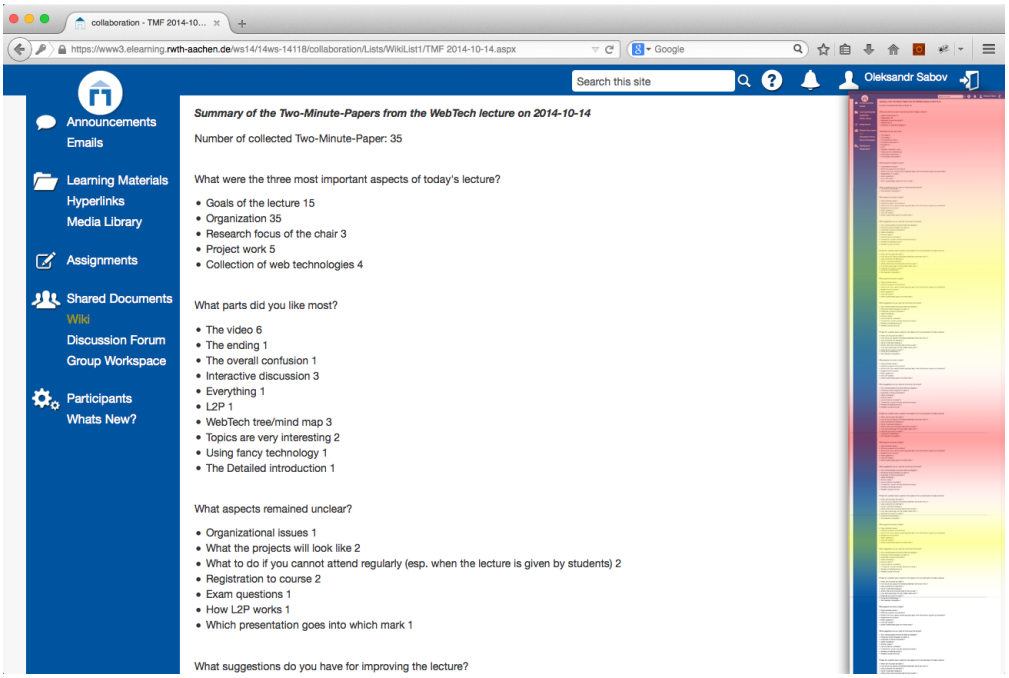
- Says nothing about the content
- Difficult to integrate with other systems

Conceptual Approach

For better support the student during the learning we should to help him during learning. It's important to enhance usual method of the obtaining information. Such a system must receive data from other similar students, and include it in the workspace of the student.

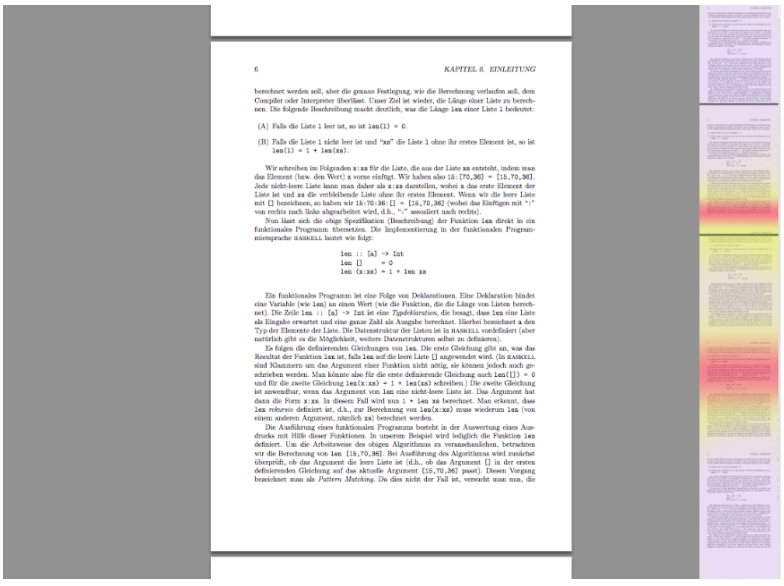
Each student is solving a problem refers to the various educational materials - books, slides, wiki pages, video lectures, etc. Dealing with this information is important experience for each student. It can be collected and shown to other students using the "Footprint" system.

The main idea is to observe how students work with different materials and process this information statistically. Each student will leave the heat trace in the book and all other students will see it. HTML wiki page look like this:

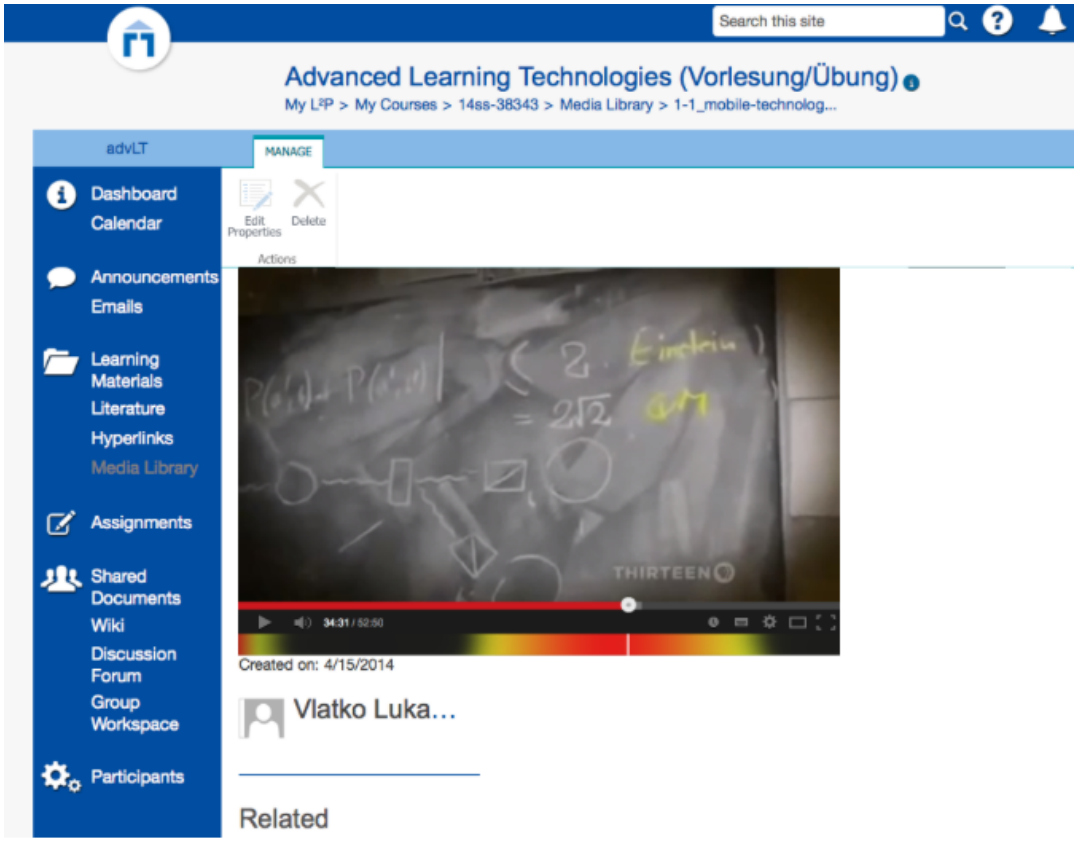


Normal scrollbar looks like a heat map, where the most popular part is red, neutral - yellow and not popular - purple. Based on this map students can begin the analysis with the most popular part. If the student does not find the information he needs, he starts to search and leaves his own trace. Every search for information in the book - it's contribution to the overall heat map. So after the exam we will get the perfect heat map of the book for students next semester.

The same for PDF book. Here's an example:



The same approach, we can apply for work with video. For example, if part of the video is not clear to the student, he rewinds this piece and see it again. On the other hand, if a student gets bored he will miss a piece of the lecture. Here is an example of how it might look:



Under the usual slider added another - heat map slider. It shows which parts of the lesson, the most interesting and which are less interesting. This solution is easy to implement, and it is the most user friendly as all users are accustomed to the normal YouTube player.

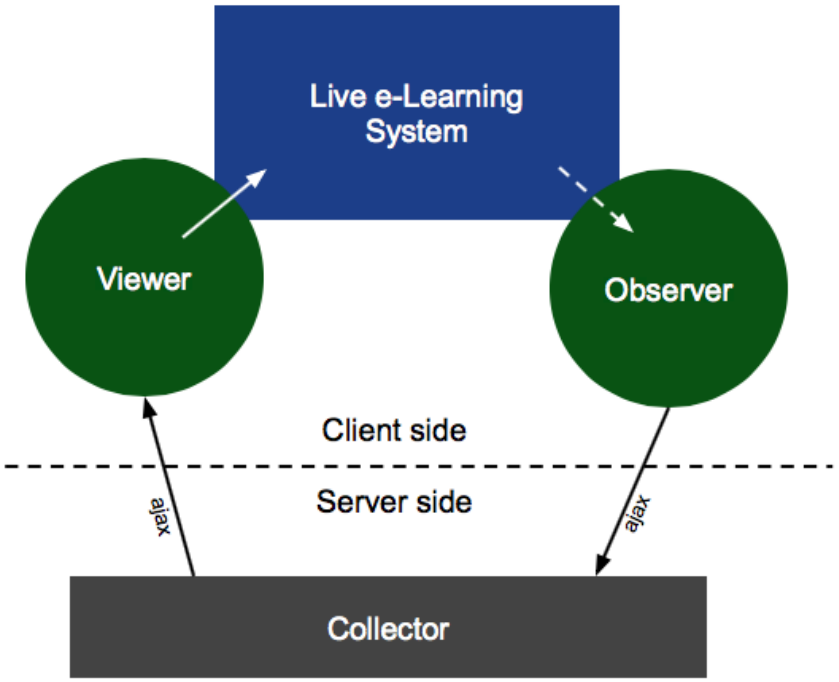
The main goal is to make the system independent. The advantages of this approach is ease integration with existing E-Learning systems. This is a very important part because connect to a live system should take place without serious interaction with the code of the E-learning system. FootPrint consists of three parts:

1. Observer
2. Viewer
3. Collector

The observer collects and processes information about the user's activity. This part will be implemented on a client side which will do all the calculations on the user's computer without impacting the server.

The viewer is the part that shows the heatmaps based on the activity of users. This part is interacts with E-Learning system as it expands functionality of the normal web page and video player.

The collector works with the data. Main task is to collect and normalize data gathered by the observer. Also, collector should return the data requested by the viewer. Also collector task is data normalization. Over time, the results will be accumulated and become not accurate and system should process it again. This diagram shows how all the parts interact:



Implementation

I'm going to use CoffeeScript on the Front and Back-End parts of the application. Viewer will use jQuery to communicate with DOM and heatmap.js to draw heatmaps. PDF viewer will based on PDF.js and custom scrollbar. To work with the player I want to add a layer that will hide the details of the interaction with the player. This approach makes it easy to add layers for different players with different APIs. The observer will also use these layers. For Back-End part I'm going to use nodeJS with Express engine and mongoDB as database. I'm going to normalize the data after each transaction. If such an approach would be too resource intensive than the normalization of the data will be held on the timer (Cron).

Also for effective work I'm going to use the unit tests (moch.js and expect.js) and automatic build system like Gulp and Bower with NPM for dependencies management. All code will be hosted on Github.

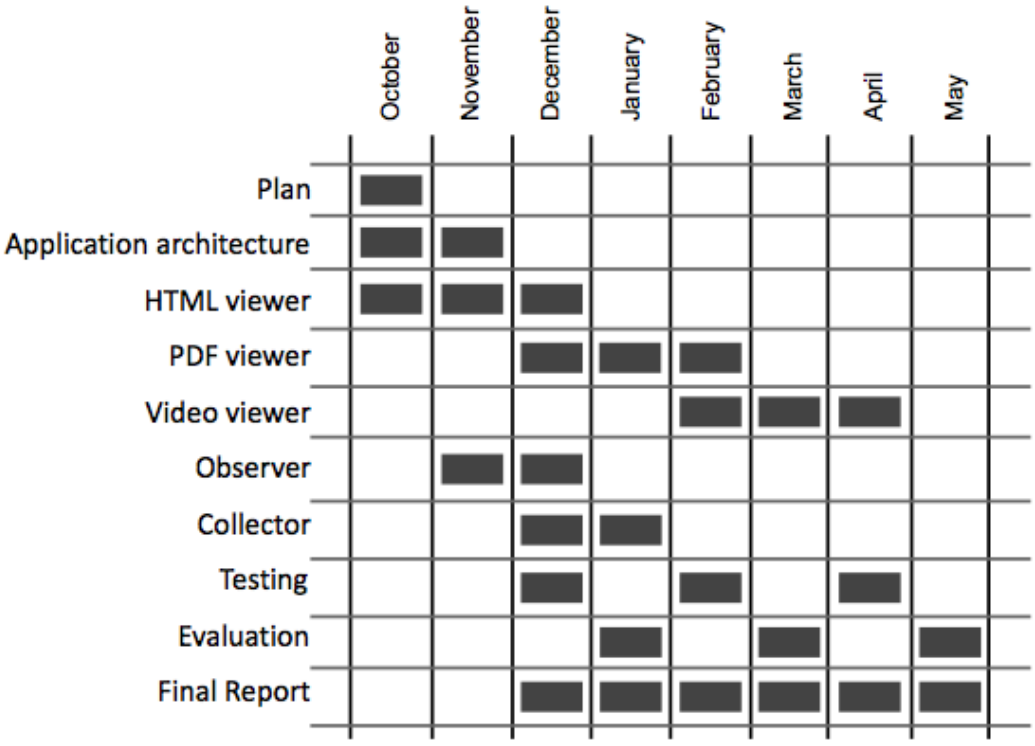
Evaluation

There are several ways to evaluate such a system:

1. Ask the user. Since the system interacts with users, we can add a small button and everyone can leave opinion about the system by answering a few questions.
2. System may be used within one academic groups. The first half will use the system, and the second is not. Such an analysis will make it clear how the system affects the performance of students, helps or not [2].
3. Poll. We can do a survey for students and receive responses from the entire group once a month.

Feedback allows us to understand is such a system useful for students or not. For best results we should use all of these types of assessments. For evaluation of the system we are going to use the live e-Learning system - RWTH Aachen L2P.

Work Plan



References

[1] Bill Meyer, Naomi Haywood, Darshan Sachdev, Sally Faraday, *"What is independent learning and what are the benefits for students?"*, London: Department for Children, Schools and Families Research Report 051, 2008

[2] Erik Duval, *"Learning Analytics for Visualization and Recommendation"*, Dept. Computer Science, Katholieke Universiteit Leuven, Proceedings of the 1st International Conference on Learning Analytics and Knowledge, pp. 14-15, 2011

[3] Mohamed Amine Chatti, Anna Lea Dyckhoff, Ulrik Schroeder, Hendrik Thüs, "A Reference Model for Learning Analytics", International Journal of Technology Enhanced Learning, pp. 6-8, 2013

[4] Janet Looney, George Siemens, *"Assessment Competency: Knowing What You Know and Learning Analytics", Promethean Thinking Deeper Research Paper No.3, pp. 5-8, 2011*