X World Conference on Computers in Education
July 2-5, 2013; Toruń, Poland

Interactive Visual Exploration of Learning Data: The Role of Teachers as Learning Analysts

Sandra Rebholz, *rebholz@md-phw.de*
University of Education Weingarten, Media Education and Visualization Group

Paul Libbrecht, *paul@cermat.org*
MLU Halle, Center for Educational Research in Mathematics and Technology

Wolfgang Müller, *mueller@md-phw.de*
University of Education Weingarten, Media Education and Visualization Group

Abstract
Modern computer-based learning tools support students in individualized, self-paced learning and provide them with immediate feedback on their performance. Using such tools as part of formative assessment practices should also inform the teachers of the students' learning activities in order to reflect and adapt their teaching strategy accordingly. Focussing on the role of teachers as *learning analysts*, we present different scenarios on how such learning tools can be applied in a university setting. Typical questions arising in these scenarios are stated and the analytical processes to answer these questions are illustrated by the example of the SMALA learning analytics platform. Based on interactive visualizations, the proposed SMA-LA log views support the teachers in getting an overview of the learning activities and let them further analyze the data according to the question of interest.

Keywords
learning analytics, visual analytics, information visualization, e-Learning, formative assessment, semi-automatic assessment

INTRODUCTION
Novel techniques and tools from the field of computer-based assessment (CBA) facilitate the assessment of individual student skills on a more regular and detailed bases compared to classical evaluations in the classroom. This allows not only for diagnostic testing in order to adapt teaching and learning processes in the sense of formative assessment. It also provides the opportunity to provide learners with adequate feedback on their performance in a more timely fashion, which is usually difficult in classical learning scenarios due to the required effort to observe and analyze students' performance in detail. Feedback, however, is considered as an indispensable element in learning to improve and accelerate student learning.

Corresponding novel techniques from CBA allow for solving exercises at a computer, recording not only the final outcome, but the complete student's learning process on the level of individual solution steps. This becomes especially important in cases where feedback is required not only on the level of effective reproduction of factual knowledge, but on the successful application of strategies and algorithmic thinking in problem-solving, such as in various fields in mathematics or its application.

The application of corresponding forms of formative CBA in general results in large amounts of log data from learner activities. In principle, this data may allow teachers to analyze a learner's performance in detail, to provide helpful feedback to individuals, and to adapt teaching in general. However, the mere amount of data from the activity logs demands for appropriate representations and techniques to facilitate the successful analysis of learner activities. For instance, summary views are required to detect symptoms such as misunderstandings or excessive successes, while sampling methods have to be supported to infer a problem seen in summaries.

A contribution of this paper is the formulation of requirements for tools supporting the analysis of learner logs based on a detailed examination of teachers' analytical processes. Furthermore, we introduce the SMALA analytical framework which is targeted to provide the corresponding functionalities in the context of the SAiL-M toolset. Finally, we present a set of tailored visualizations to support the effective analysis of complex learner log data.

The remainder of the paper is structured as follows: First, we present the theoretical background and discuss related work and approaches. Then, we introduce tools developed in the SAiL-M project (Bescherer et al., 2011) targeted to include CBA in the classroom. We also discuss briefly how this enables teachers to analyze students' performance in a new way. We present three different scenarios, which explain in more detail different motivations of teachers for analyzing learner log data, and discuss the differences in the analytical processes in detail. After this, we present elements of the analytical framework provided by the SMALA logging infrastructure including different visualization techniques. We end our paper with conclusions and remarks on open questions as a starting point for future work.

BACKGROUND AND RELATED WORK

Computers have been proposed frequently as a means to perform assessments regularly and in an effective manner. In fact, early approaches in the area of programmed instruction include this notion (Skinner, 1958). Intelligent tutoring systems (ITS) also integrate different types of assessment components as part of their inner loop (VanLehn, 2006). In both cases, the results of the assessment are being utilized to adapt the learning process but are not shown to the teachers. CBA refers to a set of different approaches for educational assessment, both in the classroom and in large-scale testing situations. Typical approaches are based on question types, with which computers can effectively interact, including scoring and score reporting, while still gathering meaningful measurement evidence (Scalise & Gifford, 2006). In practice, most of the applied techniques boil down to multiple-choice-like questionnaires, which may provide limited detail on students' conceptions and misconceptions, especially in the context of conveying knowledge beyond factual and higher-level competencies. Intelligent CBA (Bescherer et al., 2011) represents a recent approach to overcome these limitations, utilizing methods from the field of ITS to assess more complex solution processes on the level of individual solution steps. As a result, such assessments may provide much more detailed information of students' performance, raising demands in approaches to analyze the large amounts of recorded data efficiently. Different to ITS approaches, intelligent assessment does not rely on automatic assessment alone, but rather introduces a semi-automatic approach, thus causing further demand for adequate methods to support a teacher in the assessment of solution processes.

Using computer-based learning tools has the additional benefit that all sorts of digitally available learning data can be collected and analyzed. Suitable logging infrastructures are required for recording learning activities and making the recordings available for further analysis. The emerging set of specifications called Tin Can API (http://tincanapi.com/), the open-source tool sets Contextual Attention Metadata (https://sites.google.com/site/camschema/) and Learning registry (http://www.learning-registry.org/) all target at capturing learning activities and store them in a central repository. The major drawback of these solutions is the difficulty of doing statistics on the data without dedicated log viewers that are able to display detailed or statistical summaries that are expressive enough for a teacher to understand quickly the solution process or what the problems were.

Logging and processing learning data leads to the field of Learning Analytics. There have been a number of proposals for defining Learning Analytics, which to some extent take different objectives and only partially overlap (Siemens, 2008). We do connect to the field of Visual Analytics (Keim et al., 2008; Thomas & Cook, 2006), and for this reason we understand Learning Analytics as a specific focus and application area of Visual Analytics. That is, Learning Analytics relates to approaches and technologies targeted to allow for analytical reasoning facilitated by visual interfaces employed for teaching or learning. Objectives are the detection of interesting aspects and patterns in learner and learning data, building hypotheses based on these detected structures, confirming such hypotheses, drawing conclusions, and possibly communicating the results of this analytical process. In the context of this paper we will discuss how Learning Analytics relates to formative assessment, and what specific requirements can be stated on corresponding solutions from the perspective of a teacher-as-investigator. In terms of visualization techniques from the fields of information visualization and visual analytics, our approach draws from established techniques for representing tabular data, especially the Table Lens (Rao & Card, 1994; Pirolli & Rao, 1994). In addition, we apply interaction techniques from the field related to focus and context visualization (Yi et al., 2007). Furthermore, since log data is temporal in nature, our approaches also relate to the various techniques for visualizing time-dependent data (see Aigner et al., 2008 for a comprehensive overview).

INTELLIGENT ASSESSMENT TOOLS

In the context of the research project SAiL-M various interactive learning tools for the field of mathematics have been developed. Our learning tools implement the approach of intelligent assessment and use the general-purpose logging architecture SMALA (SAiL-M Architecture for Learning Analytics) for recording all semantically relevant interactions between the learners and the tools.

The SAiL-M learning tools are web-based software applications that can be accessed as learning activities from within a learning management system (LMS). Authenticated users of the LMS can use the learning tools and solve exercises interactively. The actions of the learner are analyzed and automated feedback is provided (generally detecting standard errors or standard solution paths).

The SMALA logging infrastructure provides the learning tools with the extra functionality of recording all interactions that occur between the learner and the learning tool. In order to document individual learning processes, the learning tools send all semantically relevant interactions as events to the SMALA logging service with information such as the user pseudonym, the input, and the displayed feedback. The events are stored in the SMALA database and from there the data logs get analyzed and represented by suitable log views. Authorized teachers can access these log views from the SMALA web server. Available log views include both summary views on activities and performances of the whole group of learners, and session views on step-by-step recordings of individual learning processes. We describe them below.

Evaluations of the SAiL-M tools at different universities in Baden-Wuerttemberg (Germany) confirmed acceptance and usefulness of the learning tools from the students' point of view, but also exposed the demand of teachers for appropriate analytical tools (see Libbrecht et al., 2012). Teachers requested other statistical indicators and richer summary views for getting a general overview of the students' activity and performance. In particular, they were interested in statistics on assessment results (e.g., type and number of detected problems), number of feedback requests and the level of activity per student, which were not available during the evaluation. Based on these outcomes, we have developed analytical process scenarios as a means to illustrate the integration of SMALA log views into the teaching practice and to get a better understanding of the teachers' requirements for such log views.

ANALYTICAL PROCESS SCENARIO

When integrating formative CBA tools in a classical teaching environment, the teachers typically introduce the learning tools during the lecture. The tools can then be used at home or in the lab. Based on the assessment results, the teachers can decide on appropriate adaptions of their teaching strategy. In this section, we describe illustrative scenarios involving the usage of learning analytical tools.

Scenario I: Checking the students' activity

The teacher introduces the topic of functions and relations. In order to demonstrate the concepts of injectivity and surjectivity, she interactively constructs relations by using the tool Squiggle-M, a tool to exploratively discover the properties of mappings. At the end of the session the teacher asks the students to do exercises 1 and 2 from the learning tool as homework and to perform their own explorations with the tool as they wish, with a one week delay. One day before the next session, the teacher checks whether the students have worked with the learning tool. She opens the SMALA log views for the Squiggle-M learning activity and gets a visual representation of the learners' activities. According to the diagram most of the students have solved both exercises. Their level of activity shows that they were interacting intensely with the tool. Only a few students have done only one exercise with a low level of activity. Based on her experience, she decides that a participation rate of 70% is reasonable as a proportion of high involvement.

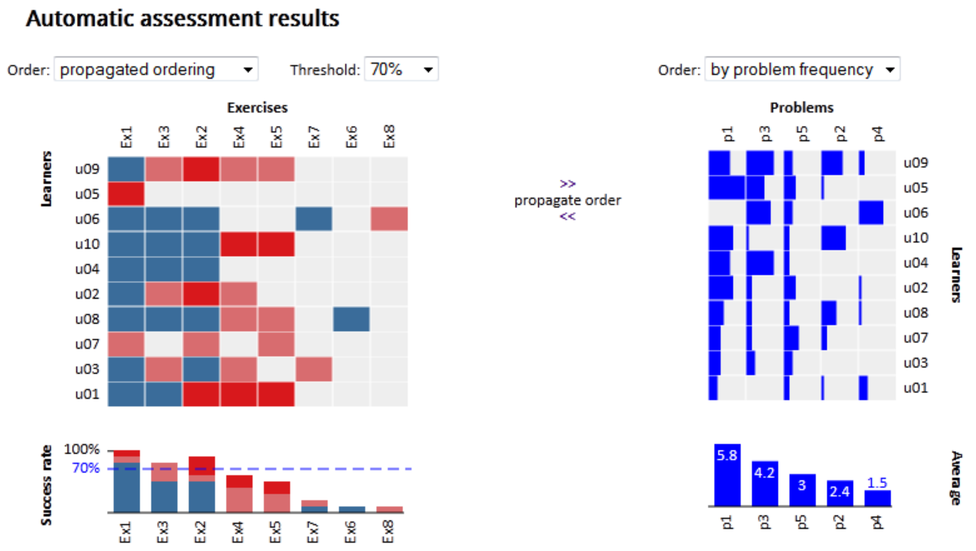

Figure 1: SMALA visualization of automatic assessment results

Scenario II: Detecting possible areas of misconception.

The teacher introduces the concept of proofs by complete induction. After explaining the underlying idea, she demonstrates how to apply the principles by proving the formula of the Gauss summation. At the end of the session, the teacher asks the students to do exercises 1 to 3 from the online training tool ComIn-M as homework. When communicating to students, she realizes that some questions clearly show misunderstandings of the process she had demonstrated to students. Using the tabular overview depicted in figure 2, she checks how successful the students were in solving the exercises. The table view is part of the SMALA log views for the ComIn-M learning activity and shows the completion state per user and exercise.

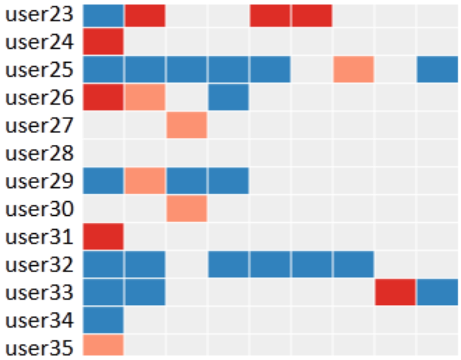

Figure 2: Detail from the “Solved exercises” view

She sees in this log view that the first exercise was much attempted and solved correctly for an amount (blue colored cells) while the second much less.

Looking at the problem types reported (Figure 3), she sees that the error message “Please give the induction hypothesis first.” has been issued more than 200 times. She can easily suppose that the induction method proof attempts are done in the wild, which often loses the learners into improper sequencing of their arguments. Thanks to hyperlinking, she can view a sample of the detailed sessions of the learners and can confirm this hypothesis.

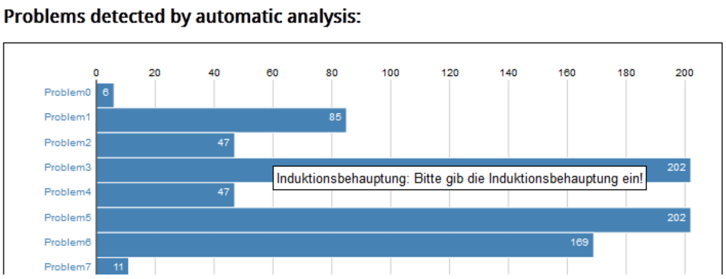

Figure 3: Detail from the “Problems detected” view

While exploring these samples, the teacher observes that a broad majority of the problems occurred in the step of finding the correct induction statement. Because the teacher considers these findings critical, she decides to explain this part of complete induction again in more detail in the coming session.

Scenario III: Providing individual feedback.

A student is doing the homework for the mathematics class and opens the learning tool ComIn-M in order to solve the exercises online. She selects the first exercise that the teacher asked the students to do. After some minor difficulties in entering the mathematical formulae, she successfully enters the base case for the proof by mathematical induction. In order to find out whether she did fine so far, she requests an automatic analysis of the current solution by pressing the “Verify” button. A green check mark appears on the screen, confirming that her intermediate solution is correct. She continues by selecting the correct assumption and then tries to figure out the induction statement. As she is not sure what to enter, she guesses a statement and requests an automatic analysis from the tool. This time, the tool marks the current solution as wrong and displays a short description of the problem that was detected. Now the student tries another solution, requests an analysis again, but again the tool reports a problem. The student is afraid that she cannot find the correct solution on her own, so she uses the “Ask Tutor” feature of the learning tool. By simply clicking the corresponding link, a dialog window opens and lets her enter a message to the tutor. When she submits the dialog, her message is sent along with a link to her SMALA logging session to the responsible tutor.

Later that day, the tutor checks her email and finds a notification that a ComIn-M user needs personal help from her. She reads the message and follows the link to the SMALA logging session. This log view shows the recorded interaction sequence between the student and the learning tool until the point of help request: it shows an easily readable overview of each the terms the student has input and all the problems that were reported by the tool. Investigating the last state of the solution process, the teacher quickly finds out that the student did not replace the index variable correctly. So she sends her advice back to the requesting student, addressing the concrete problem that she detected in the solution of the student.

The Analytical Processes

In the scenarios described above the teachers typically perform four parallel reasoning processes, which could be carried out alternating or in parallel: based on their knowledge of the domain and learning tools, they have expectations of the learners' activities; these expectations are compared to the analytical views in an explorative browsing way; this browsing leads to interpretations of the learning processes, which results in strategies being assembled to further teaching actions.

Process 1: Determine the expectations.

Based on their course plan and assignment, the teacher has expectations about the students' usage of the learning tools. Typical expectations of interest would be: this assignment should have been fully (or barely) completed since it is easy (or challenging); can we find typical problems?; this technical problem is likely to happen; or expect to see sufficient evidence in the analytics views to decide on deepening a subject or not. These expectations are constantly adjusted based on the processes below.

Process 2: Log views analysis.

Typically, teachers perform a multi-step analysis on the assessment results (Goertz, Oláh & Riggan, 2009). First, teachers look at overall scores and learning outcomes to get an overview of the general class performance. Such summaries of the automatic assessment results should highlight weaknesses both by content area and by student. Thereby, it is possible to detect common problems and difficulties. It is also possible to identify low-performing students that need special support and further guidance. In a second step, teachers perform an in-depth analysis of selected individual solutions and errors. The detailed analysis shall reveal insights into the reasons for errors. Ideally, not only the product of learning should be considered in this analysis step, but the whole process leading to the final product of learning.

Process 3: Interpret the learning processes.

The interpretation of the log views leads to an understanding of the learning process. Interpretations depend on the professional experience of the teacher and are often based on so-called "thresholds". A threshold in this context is defined as a "criteria for determining whether student[s] performance[s] require[s] an instructional response" (Goertz, Oláh & Riggan, 2009). Teachers can use thresholds as an indicator as to whether a sufficient amount of students has mastered the content covered by the assessment. This allows teachers to decide on the need for an adaptation to the teaching plan, for the classroom or for the individual student.

Process 4: Preparing the instructional response.

The outcome of the analytic process is the most important and the most challenging objective: how can the interpreted analysis be turned into action? What are the necessary measures to address the detected problems? Again, it is the experience of the practicing teacher that can help in answering these questions. In this step, research meets practice by developing concrete teaching strategies from research findings. Although it is the individual teacher who is ultimately responsible for selecting an appropriate action strategy, there are numerous sources of ideas and suggestions for finding such strategies (Altrichter & Posch, 2007).

VISUALIZATIONS

Inspired by the work of Mazza & Milani (2005), the automatic assessment results for individual users are represented in a matrix style graph. Each cell shows the completion status of one particular learner (y-axis) for one particular exercise (x-axis) coded by color. Possible values for the completion status comprise “Correct” (blue), “Wrong” (red) “Incomplete” (light red) and “Missing” (light grey). According to the example of Mazza’s GISMO system, the matrix display is supplemented by a histogram at the bottom of the matrix, showing cumulated assessment results per exercise. While the histogram provides the teacher with an overview of the total number of exercises that were solved successfully or unsuccessfully, the matrix display shows how successful and unsuccessful solutions are distributed over individual users and exercises. In order to support **interactive analytics**, the basic tabular view was enhanced by various features to sort table rows and columns and to drill down to detailed views on individual user sessions and process recordings, which are inspired from the Table Lens technique (Rao & Card, 1994; Pirolli & Rao, 1994). Sorting capabilities include the reordering of the complete matrix so that successful solutions are displayed in the upper left corner of the matrix, or reordering the complete matrix according to the values of one selected row or column. Figure 4 compares two resulting matrix permutations when ordering according to successful (left matrix) or unsuccessful solutions (right matrix).

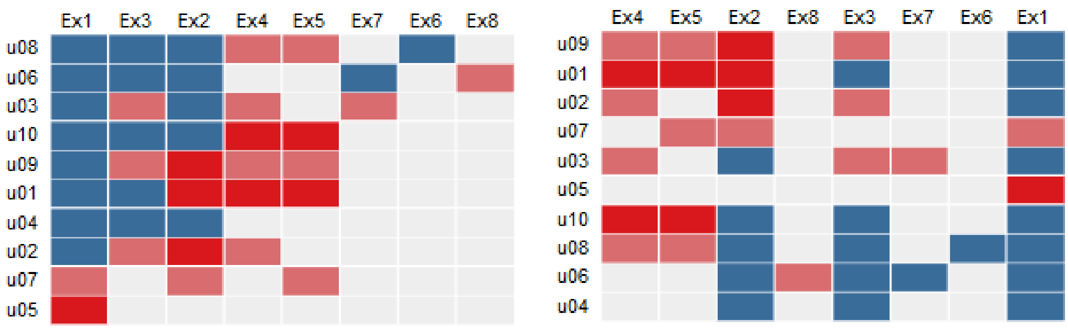

Figure 4: Different matrix permutations in the “Solved exercises” graph

As automatic assessment results recorded by SMALA do not only contain information about simple “correct” vs. “wrong” solutions, but also on the type of problems that occurred during the learning processes, the “Solved exercises” graph is accompanied by the so-called “Detected problems” graph (Figure 1). Again a tabular view is used as a main structure for displaying information. However, problem frequencies are now displayed as horizontal bars in each cell of the table. In the same way as table lens presentations (Rao & Card, 1994; Pirolli & Rao, 1994) do, individual columns (and rows) can be sorted by frequency, thus revealing patterns and relationships within the data. As in the “Solved exercises” view, a histogram is displayed below the tabular view, showing the average problem frequency of all users. Both matrices can be explored and reordered separately. However, it is also possible to propagate the row ordering from one matrix to the other. By doing so, possible correlations between successful or failed solutions and the frequency of certain problem types can be detected easily.

CONCLUSIONS

We have demonstrated by examples and task descriptions how teachers can be helped by applied learning analytics. Doing so can transform the usage of learning tools into a trustable and continuous monitor of the learners' performance. The formative computer based assessments are the key instruments to provide learning analytics that enables a witness of the learning.

Teachers employing the analytics tools need special solutions that are tuned to the learning tools. They generally follow the four processes when analyzing; their analysis is often of an exploratory nature, changing visualizations by adjustments (re-ordering, thresholding) or by hyperlinking.

Our SMALA interface provides an example of these interactive visualizations and we intend to implement and evaluate other visualization methods.

REFERENCES

Aigner, W., Miksch, S., Müller, W., Schumann, H., Tominski, Ch.(2008): Visual Methods for Analyzing Time-Oriented Data. *IEEE Transactions on Visualization and Computer Graphics*, *Vol. 14(1)*.

Altrichter, H. & Posch, P. (2007). Lehrerinnen und Lehrer erforschen ihren Unterricht (4. Auflage). Bad Heilbrunn: Julius Klinkhardt.

Bescherer, C., Herding, D., Kortenkamp, U., Müller, W., Zimmermann, M. (2011). E-Learning Tools with Intelligent Assessment and Feedback. In S. Graf et al. (Eds.): *Adaptivity and Intelligent Support in Learning Environments*. IGI Global.

Goertz, M.E., Oláh, L.N., & Riggan, M. (2009). Can interim assessments be used for instructional change? (CPRE Policy Brief R-51). Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.

Keim, D., Mansmann, F., Schneidewind, J., Thomas, J., & Ziegler, H. (2008). Visual analytics: Scope and challenges. In S. Simoff, M. Böhlen, and A. Mazeika (Eds.): *Visual Data Mining*, volume 4404 of LNCS, pp. 76–90. Springer. See http://www.springerlink.com/content/5275180h84863347/.

Libbrecht, P., Rebholz, S., Herding, D., Müller, W., & Tscheulin, F. (2012). Understanding the Learners' Actions when using Mathematics Learning Tools. In: Jeuring, J. et al. (Eds.): Proceedings CICM 2012. LNCS 7362. Springer.

Mazza, M. & Milani, C. (2005). Exploring Usage Analysis in Learning Systems: Gaining Insights From Visualisations. In: Proceedings of the workshop on Usage Analysis in Learning Systems. 12th International Conference on Artificial Intelligence in Education 2005 (pp. 65-72). Amsterdam, The Netherlands.

Pirolli, P. & Rao, R. (1996). Table lens as a tool for making sense of data. In Proceedings of the workshop on Advanced visual interfaces (pp. 67-80). ACM: New York, NY, USA.

Rao, R. & Card, S. K. (1994). The table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. In Proceedings of the ACM CHI 94 Human Factors in Computing Systems Conference 1994 (pp. 318-322), Boston, Massachusetts.

Scalise, K. & Gifford, B. (2006). Computer-Based Assessment in E-Learning: A Framework for Constructing “Intermediate Constraint” Questions and Tasks for Technology Platforms. *The Journal of Technology, Learning, and Assessment, 4(6).* http://www.dokeos.com/doc/thirdparty/Computer%20Based%20Assessment.pdf

Siemens, G. (2011). What are learning analytics? See http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/.

Skinner, B. F. (1958). Teaching machines. *Science* 128 (967-77), 137-158.

Thomas, J.J. & Cook, K. A. (2006). A visual analytics agenda. *IEEE Computer Graphics and Applications, 26(1)*, 10–13.

VanLehn, K. (2006). The Behavior of Tutoring Systems. *International Journal of Artificial Intelligence in Education, 16(3)*, 227-265.

Yi, J. S., Kang, Y. a., Stasko, J., & Jacko, J. (2007). Toward a Deeper Understanding of the Role of Interaction in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics, 13(6)*, 1224-1231.

Biographies

Sandra Rebholz is a computer scientist with a research focus on e-learning environments and modern software engineering practices. As a research assistant at the University of Education in Weingarten, she was member of the SAiL-M project team and is responsible for the development of web-based e-learning components and learning analytics tools.

Paul Libbrecht is a researcher in the Center for Educational Research in Mathematics and Technology and the development team of the Curriki.org learning resources sharing platform. He researches objects of mathematics learning on the web, including open educational resources, authoring, retrieval, and teaching analytics.

Wolfgang Müller is a professor for Media Education at the University of Education in Weingarten, where he is leading the Media Education and Visualization Group (MEVis). His research interests are in the areas of E-Learning, Learning Analytics, Human-Computer Interaction and Interactive Storytelling. He is a delegate to IFIP WG 3.3 and a co-editor of Springer Transactions on Edutainment and the Journal of Technology for Education and Learning (JTEL).