
The Student Activity Meter for Awareness and Self-reflection

Sten Govaerts
KU Leuven
Dept. of Computer Science
Celestijnenlaan 200A
3001 Heverlee, Belgium
sten.govaerts@cs.kuleuven.be

Erik Duval
KU Leuven
Dept. of Computer Science
Celestijnenlaan 200A
3001 Heverlee, Belgium
erik.duval@cs.kuleuven.be

Katrien Verbert
KU Leuven
Dept. of Computer Science
Celestijnenlaan 200A
3001 Heverlee, Belgium
katrien.verbert@cs.kuleuven.be

Abelardo Pardo
Universidad Carlos III de
Madrid
Avenida Universidad 30
E-28911, Leganés, (Madrid)
abelardo.pardo@uc3m.es

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI'12, May 5–10, 2012, Austin, Texas, USA.

Copyright 2012 ACM 978-1-4503-1016-1/12/05...\$10.00.

Abstract

Visualization of user actions can be used in Technology Enhanced Learning to increase awareness for learners and teachers and to support self-reflection. In this paper, we present our Student Activity Meter that visualizes learner actions. We present four design iterations and results of both quantitative and qualitative evaluation studies in real-world settings that assess the usability, use and usefulness of different visualizations. Results indicate that our tool is useful for a variety of teacher and learner needs, including awareness of time spent and resource use. Tools like SAM can also be deployed in other settings that require awareness and self-reflection, e.g. in personal informatics and health monitoring, where motivated users will value the flexible mechanisms to analyze trending data.

Author Keywords

visualization; awareness; self-reflection; monitoring; teaching; evaluation

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces – User-centered design;

Introduction

In a physical classroom, learners interact directly with teachers (even though orchestration can often be problematic [4]) and other learners. In a digital environment, learners often struggle to understand whether their work is of the same level as that of their peers or whether their work is in-line with the expectation of the teacher [1]. Teachers experience similar problems: it can be difficult to track learner progress or to detect learners experiencing difficulties.

In this paper, we apply visualization techniques to enable understanding and discovery of patterns [5] from data that is tracked in learning environments. The overall goal of our Student Activity Meter (SAM) application (see Fig. 1) is to assist both teachers and learners with self-reflection and awareness of what and how learners are doing. This can be especially useful in self-regulated learning [37], where learners are in control of their own learning. We applied a design-based research methodology [18] to evaluate the usefulness and usability of different visualization techniques to support these goals. These evaluations were conducted in real-world courses with students and teachers in four iterations over a period of 24 months.

The paper is organized as follows: in the next section, we give a brief overview of related research. Then, objectives for both learners and teachers are discussed. Next, we discuss the design iterations and evaluation results. Conclusions and remarks on future work are discussed in the last section.

Related Work

Visualization techniques have been researched and applied in several domains [14], including finance [35] and the

automotive industry [22]. In this paper, we focus on the Technology Enhanced Learning (TEL) domain. Two categories of visualization systems are used to support self-reflection and awareness among learners [26]. A first category includes systems that improve awareness of actions taken on resources. A second class supports aggregating data into a set of high-level indicators, e.g. analyzing participation rates based on word counts and message reply delays. Our tool, SAM, provides visual analysis of time learners spent on learning activities and document use statistics. Hence, it belongs to the second category.

Several available tools belong to the first category. Hardy et al. [12] visualize the path taken through resources. These visualizations are in essence directed graphs, with nodes representing pages and edges representing access between pages. GISMO [16] visualizes accesses to a course and its resources. Among others, the application relies on a simple matrix visualization of student names and resource names to represent resource accesses.

CAMera [21] is an example of an application that belongs to the second category. CAMera analyzes user activities and provides simple metrics of events, such as mouse clicks. The system relies on Contextualized Attention Metadata (CAM) [33] that captures user interactions with tools and resources. SAM also relies on CAM data, but focuses on higher-level indicators. For instance, SAM analyzes user events to extract indications of the time learners have spent on learning activities. Moodog [34] visualizes metrics of activity logs in Moodle¹ using bar charts. SAM provides more fine-grained time spent and uses a wider range of more advanced visualization

¹Moodle, <http://www.moodle.org>

techniques (such as parallel coordinates) enabling more detailed analysis.

Most visualization research for TEL only discusses how and for which purpose the tools can be used (e.g. [21]). Only a few in-depth studies with real teachers and learners (e.g. [13]) have been conducted that assess the usefulness of visualizations. In this paper, we elaborate such research and present SAM's iterations and evaluations with teachers and students. The major contribution lies in a presentation of insights into how and why teachers and learners use visualizations to improve awareness and self-reflection in real-life scenarios.

Objectives

SAM visualizes data tracked in learning environments. Depending on the level of detail in the data, different metrics are provided, like basic time spent and resource use or forum view and post actions.

Both teachers and learners are target user groups. SAM assists *teachers* with awareness of what and how students are doing. More specifically, SAM can assist teachers to detect students doing well and at risk. SAM uses time spent and resource use metrics [15]. Teachers gain insight into the time learners are spending on course activities and can compare these indicators with their estimates. Resource use can provide insights on how and how often the documents and tools are used by students. It can also enable discovery of resources external to the course.

Metrics can be added depending on the data detail to increase understanding of specific tasks, e.g. the compiling error count or number of forum contributions.

SAM also supports self-reflection and awareness for *learners*. SAM visualizes how long learners have been

working on course activities in comparison to their peers. Discovery of averages and trends is possible, as well. In addition, resource use indicators provide information about learning material used by other students. This can be especially useful in self-regulated learning [37].

Design iterations and case studies

To design and develop SAM, we applied the design-based research methodology [18] that relies on rapid prototyping to evaluate ideas in frequent short cycles. SAM evolved during four iterations over 24 months. Fig. 1 shows the latest version of SAM (Iteration 4) and is used below to illustrate the design changes through all iterations.

Table 1 presents an overview of all iterations and their main evaluation outcomes. All evaluations have been conducted in real-world settings using real courses in a variety of topics and their teachers and students. SAM is fully implemented in Flash, enabling web and desktop availability and relies on CAM for data interoperability.

Iteration 1

In this section, we present the first version of SAM. More details and screenshots of iteration 1 are available in [11].

Design and Implementation

SAM has been designed and implemented using paper mockups – which were refined based on feedback from colleagues. Our visualizations provide coordinated views [32] over the data.

The line chart (label 4 in Fig. 1) enables finding trends [27]. Every line represents one student, connecting all the timestamps when he was working with the cumulative amount of time spent. The inclination of the line shows the effort of the student. A steep line means an intensive working period. A flat line shows inactivity. The

iteration	demographics	period	methodology	evaluation goal	design adaptations	SUS	evaluation results	
							negative	positive
1	12 CS students	April - May 2010	task-based interview with think-aloud, SUS and MSDT	usability, user satisfaction and preliminary usefulness	first implementation (see "Iteration 1")	73.33	some small usability improvements can be made	learnability is high low error rate user satisfaction and usability are decent preliminary usefulness are positive line chart and parallel coordinates most useful
2	19 CGIAR, CS & law teachers & TA's	Sep 2010	online survey with Likert items, open & yes/no questions	assessing teacher needs, use and usefulness	help function		resource recommendations not useful	SAM provides awareness all visualizations useful found many uses for SAM 90% want to continue using SAM
3	12 LAK course participants (teachers & visualization experts)	Mar - April 2011	online survey with Likert items, open & yes/no questions	assessing teacher needs, expert feedback, use and usefulness	re-orderable parallel coordinates with histograms		most addressed needs are indecisive	SAM provides awareness SAM assists with providing feedback found many uses for SAM resource recommendations can be useful 66% want to continue using SAM
4	11 CS teachers & TA's	Sep 2011	structured face-to-face interview with tasks & open questions	use, usefulness, user satisfaction	filter and search functionality, icons per resource type, vertical axes zooming in line chart, adding/removing parallel coordinates' axes	71.36	conflicting visions of students doing well or at risk	SAM provides time spending overview SAM provides better course overview parallel coordinates assist with detecting problems many new insights gained by teachers line chart and parallel coordinates most useful found many uses for SAM 100% want to continue using SAM

Table 1: Overview table of SAM's design iterations and main evaluation outcomes.

selected student (in red) worked intensively immediately at the start, then stagnated a bit and stopped working by the end of September. Because the display can get quite dense, one can zoom on date (label 7). Vertical zoom (8), search (9) and filtering (10) were unavailable in iteration 1. The chart allows to compare learners and find out on what activities or resources they spent their time.

Parallel coordinates (label 5) are a common way to visualize high dimensional data in limited space and allow exploration of trends [24]. The vertical axes represent different metrics. The metrics differ based on the available data. The metrics for time spent and resource use available in iteration 1, 2 and 3 are the total time

spent on the course, the average time spent on a resource, the number of resources used and the median of time of day when students work. Each student is shown as a polyline. The yellow polyline shows the average student (A, in yellow). The selected student (S, in red) used 250 resources, on average for 4 minutes, works longer than the average student and has the 2nd highest number of programming errors (this metric was unavailable in iteration 1). Parallel coordinates are useful for data exploration and can allow an overview of the tendencies and positioning of students [24].

The bar chart (label 6) shows the student distribution in a histogram for the total time spent and the resources used.

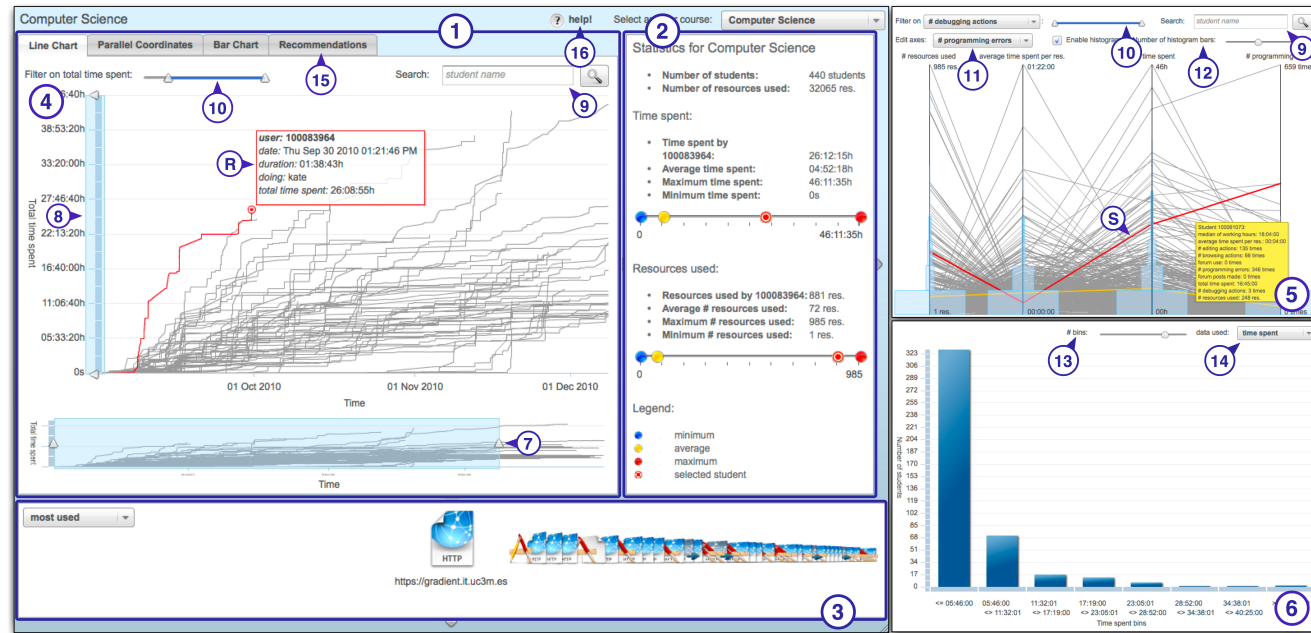


Figure 1: The user interface with three visualizations.

This allows student grouping and provides a visual impression of the distribution. One can drill down in the data by clicking on a specific bar. The number of bins (13) and the metric (14) can be changed.

The statistics of time spent and document use are shown in box 2. A graphical view is presented with color-coding of the minimum, maximum and average time spent and documents used. Upon selection of a user in the visualizations, his statistics are shown.

The recommendation pane (label 3) enables navigation through used resources. In iteration 1, different icons

based on resource type were not provided. The 'Recommendations' tab (Fig. 1 label 15) contains an animated tag cloud of the recommendations. We opted for a tag cloud, because the information-seeking task is not clearly defined and exploratory [25].

Evaluation with students

Evaluation Data

For this iteration, we implemented an online tracking tool using Twitter to log time spent and resource use [11]. The tracked data is preprocessed and exposed by the online tool to improve visualization performance.

Demographics and Evaluation Setup

The usability and user satisfaction was evaluated with 12 computer science students (10 male and 2 female, from 22 to 25 years old) at the Katholieke Universiteit Leuven taking a HCI course in April-May 2010.

Two evaluation sessions were conducted. After a week of time tracking, a task-based interview was conducted using the think-aloud protocol [29] to provide more insight in the student's understanding of the visualizations and evaluate the learnability, errors and efficiency [17]. After a month of use, user satisfaction was evaluated with the System Usability Scale (SUS) [8, 28]. In addition, we applied the Microsoft Desirability Toolkit (MSDT) [7] to gain more detailed and descriptive user satisfaction information.

Evaluation results

Learnability, errors and efficiency were evaluated with 7 students. Generally, the students could explain well what they saw on the visualizations. Two students required our explanation that every line in the parallel coordinates and line chart represents a student. Two students had trouble finding info boxes in the line chart (the red box with label R in 1, Fig. 1), because it requires hovering over a specific point.

Overall, the visualizations themselves are clear – only two people required an explanation that every line represented one student. In iteration 2, we added documentation for first-time users to remedy this. The operation of the line chart zoom tool requires more information, in iteration 2 a tooltip was added. The efficiency was not timed. Although overall no real problems were uncovered, improving the line selection would be good.

The User Satisfaction was measured with nine students, of which five were veterans of the first round. The



Figure 2: A word cloud based on the frequency of the selected MSDT adjectives.

average SUS score is 73.33 with a standard deviation of 9.35. This rates above the average SUS score (69.69) of 206 studies [6]. The SUS score of three students is above 80 and two are lower than 70. From the frequency of the MSDT results we created the word cloud in Fig. 2.

'Usable', 'Useful', 'Understandable' and 'Organized' are the most selected words. The words 'Relevant', 'Effective', 'Meaningful', 'Straightforward', 'Easy to use' and 'Intuitive' indicate a high level of usability. These results are in-line with the SUS results. Three students selected 'Complex'. Two participants indicated that this was just their initial reaction. For the other student complexity was positive – meaning 'advanced' or 'evolved' and not 'difficult to use'.

We applied verbal protocol analysis to the top five choices of the students: 42 words are positive (93%) and 3 words are negative (7%). The negative words are 'overwhelming', 'complex' and 'cluttered'. Both results indicate a decent user satisfaction, but first-time users may need some help.

The usefulness of SAM was assessed by asking why it would be useful to them. Six students use SAM to see how much the others work on the course for comparison and three are mainly interested in their own data for reporting. This result is very much in line with our objectives. Four students liked the line chart most, three selected the parallel coordinates and two the bar chart. Initially, we included three visualizations because we did not know what people would prefer. It is a bit surprising that the parallel coordinates score so high, because this visualization is harder to grasp at first [24]. The students who voted for this, said they like it for the clear overview and to see trends.

Conclusion – Learnability is high. The number of errors is low and no unrecoverable errors are encountered. The SUS test with an average score of 73 and the MSDT indicate a decent user satisfaction and usability. Preliminary results of how students use SAM are positive and in line with the objectives.

Iteration 2

After measuring the usability with students, we evaluated whether SAM can assist teachers with their teaching tasks. The teachers of CGIAR², an agricultural research group, teach many agroforestry and science courses on Moodle, which logging features provide us with tracking data allowing deployment of SAM over multiple courses and evaluation with multiple teachers.

Design and Implementation

In this iteration, we introduced a help function (label 16 on Fig. 1) to assist first-time users. In addition, some

bugs were fixed and Google Analytics³ was integrated to monitor users during the evaluation.

Evaluation with CGIAR Teachers

Evaluation Data

Scripts were developed to transform Moodle log events into time spent and resource use data. A sliding time window algorithm is used that groups events based on resources and tools used. The algorithm uses different thresholds to correct user inactivity (e.g. when a user leaves the computer). The scripts run at fixed times to provide SAM⁴ with faster performance and recent data.

Demographics and Evaluation Setup

In total, 20 persons participated of which only 3 CGIAR teachers. To increase participation, other teachers contributed: 2 law, 8 Computer Science (CS) teachers and 7 CS teaching assistants (TA). They were not involved in the CGIAR courses, but as teachers they could still provide valuable feedback. There were 6 females and 14 males, between 24 to 66 years old. Ten teach more than 5 years.

An online survey was used, because the teachers are spread over 3 continents. The survey inquired about the importance of teacher issues, with which teacher issues SAM assists and the usefulness of SAM through a detailed study with Likert items, eight open questions and multiple yes/no-questions on usage intentions. The qualitative data provided us with more details and better understanding of the quantitative data. We designed our own survey to explore more specific topics on how teachers use SAM and it assists them.

²CGIAR – Consultative Group on International Agricultural Research, <http://cgiar.org>

³Google Analytics, <http://analytics.google.com>

⁴The tool is available at <http://ariadne.cs.kuleuven.be/monitorwidget-cgiar>

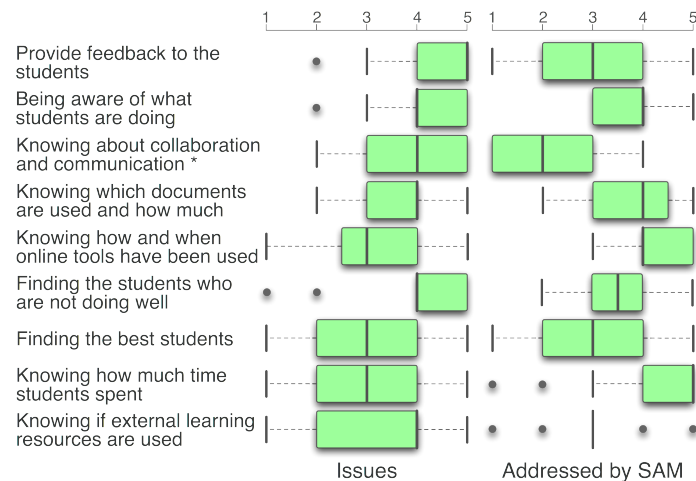


Figure 3: Box plots of the Likert scale analysis of teaching issues for the CGIAR evaluation.

Evaluation results

The respondents first rated 9 5-scale Likert items about teaching related needs and issues, based on the teacher objectives. Afterwards, they rated whether the same issues could be addressed by SAM. This way, the real needs can be related to how they perceive SAM addresses these problems.

Teaching needs and issues – Figure 3 shows the Likert scales ('1 - not relevant' to '5 - very important') on teaching needs and how well SAM addresses these. The most important issue for teachers is providing feedback. 'Knowing about collaboration and communication' (the 3rd row with *) is not addressed by SAM, but is added to check a possible bias. The difference between teachers and TA's has been statistically tested with Mann-Whitney's U tests. Giving feedback is more

important for teachers than for TA's (mean ranks of teachers = 12.19 and TA's = 7.36; $U = 23.5$, $Z = -2.16$, $p = 0.02$, $r = 0.48$). TA's find collaboration significantly less important than teachers (mean ranks of teachers = 12.73 and TA's = 6.36; $U = 16.5$, $Z = -2.40$, $p = 0.01$, $r = 0.54$). The results show that the awareness objective (through 'provide feedback' and 'being aware of what students are doing') and resource use (through 'knowing document use' and 'knowing external resources use') are important. Time tracking is less important.

The addressed issues are assessed with the Likert items ('1 - impossible' to '5 - definitely') in Figure 3. The collaboration issue is rated the lowest as expected. However, the teachers do not fully agree (high range and IQR). The highest rated is 'knowing how much time students spent' and 'Awareness of what students are doing'. Finding students in trouble and the best students is also rated rather low. Awareness of resource use is mostly met, but can be improved by differentiating external resources (the external resource use issue is indecisive). The range is quite high for most issues, which can be related to the high-level issue descriptions that maybe require more work to relate to SAM's functionalities. SAM partially meets the awareness objective through 'Awareness of what students are doing', but can be improved. Some participants argued that time spent and resource usage is insufficient and that in most cases quantitative data does not demonstrate that students learned.

The usefulness of each visualization is presented in Figure 4 using Likert items ('1 - not useful at all' to '5 - very useful'). All contribute to the teacher needs (all medians and modes = 4). The line chart scores highest, then the bar chart and then the parallel coordinates. Two

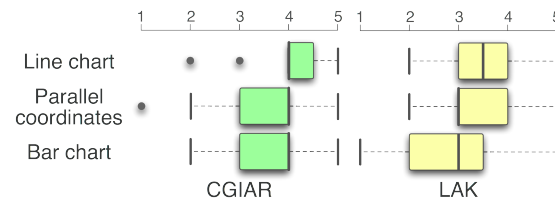


Figure 4: Box plot of the Likert scale analysis of the usefulness of the visualizations of the CGIAR and LAK case study.

participants rated parallel coordinates as useless. This result is quite in line with iteration 1.

The achievable tasks of each visualization were inquired with open questions. The time spent patterns are important for the teachers to control the workload during the course and to see when students start working. The line chart use is mostly related to awareness, finding students in trouble and time tracking. The parallel coordinates are mostly used to get a course overview, to detect students who need assistance and to compare with the average student. The bar chart is used to check the student distribution for a course overview and to find outliers. Only 5 teachers find the current resource recommendations useful. Teachers want to understand the resource use and are not interested in recommendations as such. The tag cloud is not always clear and some prefer a new list view. 14 want to give SAM to their students.

Suggestions for improvement include drilling down in the data (e.g. to obtain a student list from a bin in the bar chart) and more search capabilities (e.g. find a specific student). The teachers also requested more metrics (e.g. the number of times a student works on the course) in the bar chart and parallel coordinates. The latter would also benefit from configurable axes (see iteration 3).

The gained insights from using SAM were: time spent is a good indicator of effort (3p.), better understanding of the course workload and usage (3p.), understanding the student distributions (3p.) and finding outliers (3p.). 18 out of 20 would like to use it in their own courses.

The participants really used SAM during the evaluation. Google Analytics tracked 48 visits, spending 40 minutes on average. When removing two outliers, the average time spent is 29 minutes.

Conclusion – The most important teacher issue is providing feedback to students. Awareness and resource use is also rated as very important. The usefulness of SAM to meet the awareness objective is rated as high. Many different uses for SAM have been found from detecting students at risk to gaining a course overview. All visualizations are perceived as useful. Teachers would like more statistics, differentiate external resources and drilling down in the data.

Iteration 3

In this iteration, we extended the parallel coordinates and administered an almost identical survey to the one in the previous iteration. For this evaluation we wanted to get expert feedback and see how SAM would operate in a large course. SAM was deployed in an open online course on Learning and Knowledge Analytics (LAK)⁵ – an emerging research domain in TEL that focuses on better measurement, analysis, visualization and reporting of data about learners [2]. More details and screenshots of iteration 2 and 3 are available in [10].

⁵LAK course, <http://scope.bccampus.ca/course/view.php?id=365>

Design and Implementation

The *parallel coordinates* allow re-ordering of the axes through drag-and-drop for better metrics comparison. To cope with the line density better, configurable histograms (12) are added to the axes.

Evaluation with the Learning and Knowledge Analytics course participants

Evaluation Data

Moodle and its forums were used for collaboration⁶ and its logging data is imported identical to Iteration 2.

Demographics and Evaluation Setup

270 participants registered for the course. They are researchers active in the learning analytics field and/or interested teachers. Evaluating SAM in this diverse test bed triggered interesting discussions⁷. Twelve people, between 27 and 62 years old, participated. Six are teaching and five have done so for more than 10 years.

We again used an online survey but with two paths: one for all users and one for teachers. The path for teachers is identical to the CGIAR case study to allow comparison. Some additional questions were added.

Evaluation Results

Teaching issues – Six teachers rated the relevancy of the same teaching issues as in the CGIAR evaluation. Figure 5 shows that LAK teachers perceive teaching issues slightly different than the CGIAR teachers. ‘Providing feedback’ is again the most important issue. Both LAK and CGIAR teachers want to understand the document use. The main differences between LAK and CGIAR teachers are: LAK

⁶The tool is available at <http://ariadne.cs.kuleuven.be/monitorwidget-lak11>

⁷<http://scope.bccampus.ca/mod/forum/discuss.php?d=16525>

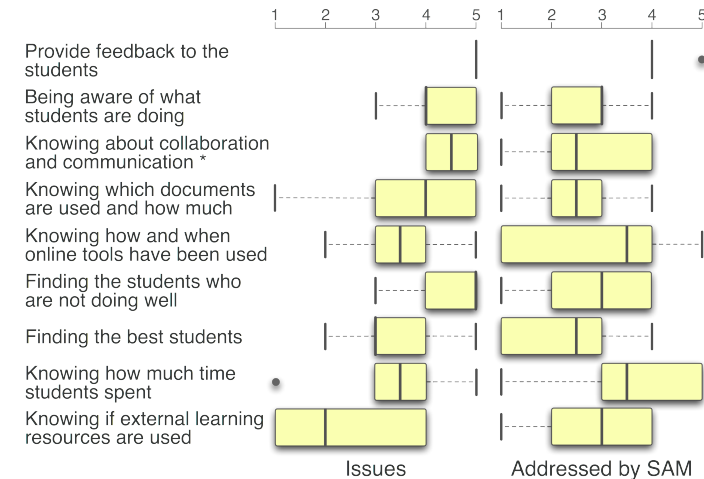


Figure 5: Box plots of the Likert scale analysis of teaching issues for the LAK evaluation

teachers rate finding students at risk higher and finding good students lower, online tool use is not so interesting for LAK teachers and collaboration is more important. Awareness is also rated high. Comparing with the objectives, awareness and resource use is again the most important.

The *addressed issues* in Figure 5 depicts that LAK teachers are more convinced that SAM addresses the ‘providing feedback to students’ issue. This might be related to their visualization expertise. The time spent issue is scoring second best, but it is less convincing than in the CGIAR case. On most other issues, the results are inconclusive (most medians about 3).

The *usefulness of each visualization* was assessed by both teachers and learners. When asked to assess the

contribution of each visualization, answers were inconclusive (all mode & median around 3), see Figure 4 (right side). There is no statistical evidence that teachers and learners rate the visualizations differently.

The achievable tasks of each visualization was assessed identical to the CGIAR case and gives better insights on the use of SAM. The line chart can be used by teachers to detect the participation intensity and tendency, to verify the classroom activity and to learn about 'chronological course dwell time'. Learners use it for comparison with his peers. The parallel coordinates are used for comparison with the rest of the class, to measure progress and to increase motivation. A learner used the bar chart to find the group where he fits in. The bar chart was perceived as a bit redundant by one teacher, maybe due to the addition of the histograms on the parallel coordinates.

Seven out of nine users found the recommendations useful. Eight users wanted to continue using SAM and four were unsure. People liked SAM for its simplicity and the multitude they can observe, and the accurate, detailed time information and the insight and motivation it provides. Although results of Figure 5 and 4 are hard to interpret, we can derive from the open questions and willingness to continue using SAM that most LAK users, perceive it as useful for different tasks. The study is affected by the incomplete user activity data of some users, who studied outside of Moodle.

Suggestions for improvement – We asked to prioritize visualization ideas from the CGIAR evaluation. From most desired to least, these are the results: (1) visualization on student content creation, (2) used resource types visualization, (3) detailed statistics of used resources, (4) comparison of the actions of two students and (5) detailed student information.

Conclusion – The LAK evaluation results confirm some of the CGIAR findings. The most important teaching issue is again 'providing feedback to students' and 'student awareness' is rated high again. The LAK teachers are very interested in finding students who are not doing well. Both LAK and CGIAR teachers want to understand document use. LAK teachers are more convinced that SAM enables them to provide feedback. In general, they are indecisive on the other issues. From the open questions, we learn that LAK participants use SAM for similar tasks as CGIAR teachers, e.g. finding outliers. They are also interested in student content creation visualizations.

Iteration 4

SAM's usefulness was assessed with teachers of a computer science course taking place last year, allowing a retrospective and full course view. The more detailed tracking data allows extra metrics and the course has a large and diverse teaching team.

Design and Implementation

In iteration 4, filtering and searching in the visualizations is implemented, as requested by teachers in the CGIAR and LAK case, to facilitate exploration of larger data sets. Adding zooming and search features also strengthens the Visual Information Seeking Mantra [23]. The following extensions are implemented:

The recommendation pane (3 in fig. 1) uses different icons for each resource type (e.g. URLs) and a specific icon for external resources to address teacher remarks.

The line chart provides zooming on the vertical time spent axis (8) and, a search (9) and filter (10) function.

The parallel coordinates have functionalities to configure more metrics by adding and removing axes (11) as

requested by the teachers. Filtering (10) and searching (9) is also integrated.

Evaluation with CS teachers

Evaluation Data

Data was collected during a Systems Architecture course at the University Carlos III of Madrid. Most of the course work is done on a development platform. To overcome potential configuration problems, a virtual machine with all the development tools is made available for students. This environment contains a tracking tool that records most of the user actions. Students are informed of the presence of this tool, the data that is being recorded and the procedure to disable it.

An individual activity report is created for each student. The data set is converted to CAM and preprocessed for better performance of SAM⁸. The richer data allows the creation of a course specific metrics set: URLs opened in the browser, debugging actions, editor use, programming errors, course forum accesses, and forum posts.

Demographics and Evaluation Setup

The evaluation was conducted with 2 professors, 5 assistant professors (AP), 1 TA, 1 course planner and 1 TA from a related course. All of them have been teaching this course for 3 years, except one AP with two years, and two other APs who were new this year.

We conducted a qualitative study with 11 teachers after the course was finished, using a structured face-to-face interview with a fixed questionnaire of 25 open questions and tasks to be able to ask them to expand on answers. Before the interview started, the interviewer presented SAM in a 5 min showcase.

⁸The tool is available at <http://ariadne.cs.kuleuven.be/monitorwidget-chi>

Evaluation Results

In total, the interview consisted of 6 tasks to get acquainted with SAM (not discussed here), 6 tasks to assess teacher needs and issues that were identified in iteration 2 and 3, and 13 questions regarding use, usefulness and preferences. We first present the 6 task results on the parallel coordinates, student feedback and time spent and then discuss the open questions.

The parallel coordinates were evaluated with 3 tasks to understand if teachers can collect findings. First, they were asked to analyze if students wrote on the forum often. Then they had to assess if the contributing students had problems with programming. For both tasks, all participants, except the course planner, were able to answer these questions based on the provided visualizations. To encourage the teachers to explore the parallel coordinates further, they were asked to relate other metrics and the forum to their experiences. All teachers could relate findings back to the course.

Teaching issues – Providing student feedback is one of the most important teacher issues, as identified by CGIAR and LAK teachers. All teachers could identify students doing well and students at risk. They posed many different, sometimes conflicting, visions of which metrics and values are to be considered. This shows an opportunity to use SAM to explore potential propositions and analyze these further. Teachers could find indications for potential students doing well and at risk on both individual level as well as a general level by setting thresholds for the metrics. The thresholds seem to be personal for each teacher.

Time spent and resource use are the two main metrics employed in SAM [11]. All teachers could estimate the time spent of the student majority and confirmed that SAM enables them to better understand how students

spend time. Six used the parallel coordinates (of which two employ the histograms) and five the line chart.

The achievable tasks and usefulness were assessed with 7 open questions. All teachers agreed that SAM provides them with a better overview on the course. Two participants noted that the line chart help to grasp an overview. All teachers agreed that SAM helps them to understand student time spent and everyone wants to use SAM during the course this year. Main benefits are: detecting students that require assistance (5 p.), course planning (3 p.) and monitor tool and resource use (3 p.). All agreed that SAM assists them in their teaching tasks. The resource recommendations are useful according to 10 teachers – mainly to know the documents and tools students use most and to know which documents need improvement or advocating with the students. Nine persons would have their students use SAM.

Privacy is a serious concern and mentioned by users in previous studies [11, 10]. In the evaluation setup, only a student number is visible to obfuscate the identity. When asked if full names would increase SAM's usefulness, ten users said that they wanted to see the names, e.g. for contacting them for feedback.

Suggestions for improvements came up during the interview and were also discussed in 5 questions. When asked what they liked most about SAM, five persons answered the line chart, because it gives them an immediate overview, summarizes time and allows discovery of outliers. Four preferred the parallel coordinates, although one mentioned it is overwhelming at first. Answers on what they like least are: sometimes too many lines overlap (3 p.) and selecting a student should be easier (3 p.). Techniques to order axes more optimally [9, 30] and to facilitate pattern discovery [36]

are available and can be integrated in the future. A list of common suggested improvements includes: visualizations for student groups and metrics over groups (7 p.), visualize metrics over time (4 p.), add a line for the 'expected student' so teacher can compare with their expectations of what a student should do (3 p.), add all metrics to the stats window (3 p.) and add filters to limit the resource recommendations (2 p.). Two more interesting suggestions were made to make SAM more social: a way to share patterns with colleagues and an alert system to warn teachers when metrics pass a threshold. When asked what extra metrics would be useful, these were most common: student marks (5 p.), group-based metrics (2 p.), more metrics on the use of different tools (2 p.), resource metrics only for those provided by the teachers (2 p.).

The user satisfaction was conducted after the interview with a SUS survey [8] to compare it with iteration 1. The average SUS score for the study is 71.36, which rates above the average SUS score (69.69) [6]. This is comparable to iteration 1 (73.33), although extra advanced features have been added (e.g. filtering and configurable parallel coordinates axes). Based on a 5-point Likert item on whether they would recommend SAM to their friends or colleagues ('1. very unlikely' to '5. very likely'), users would advocate SAM ($\bar{x} = Mo = Md = 4, R = 2, IQR = 1$). This also relates to whether they want to use the system frequently, a question in SUS, which was also rated high.

We did not find significant differences between the teachers and TA's. Some answers relate closely to the participant's teaching tasks, for example TA's want to see only 'their' students and use SAM to check up on students before an appointment.

Conclusion – Every participant confirmed that SAM helps them to get an overview of student time spent and a better overview on the course in general. All teachers can find students who are doing well or are at risk. However, they posed many, sometimes conflicting, visions on exactly which students are doing well or at risk. This can be related to their specific background, course involvement and requirements. Everyone who taught the 2010 course can relate findings back to the course. The parallel coordinates can assist teachers with detecting anomalies, patterns and potentially problems earlier on. Almost all users want to view student names to enable easier feedback, although some argue that for getting an overview and analyzing on class level anonymized data is enough. The line chart and parallel coordinates are liked most by users. The bar chart provides quite similar functionality as the histograms on the parallel coordinates. Solutions for analyzing groups and easier student selection are wanted. From the plethora of new insights gained in this one-hour session and the fact that all teachers want to continue using it in their course, we can deduct that SAM is useful to teachers in a real-life course for much more tasks than we originally envisioned.

Conclusion and future work

This paper presents the iterative design and evaluation process of SAM, a visualization tool for awareness and self-reflection for teachers and learners. Each iteration is evaluated with students, teachers and visualization experts in real-world settings (see Table 1).

First, SAM's usability was evaluated with students and preliminary indicators of its usefulness for students were found. Next, teachers from CGIAR were inquired about their needs and whether SAM could meet those needs. The most important teacher need that SAM addresses is

to provide feedback to students. Many uses were mentioned (e.g. finding outlier students), which indicates a perceived usefulness and contributes to better awareness. From participants in a learning analytics course, we learned that they have similar teaching needs, but they are indecisive whether SAM assists them with these needs. They would use SAM for similar reasons as the CGIAR teachers. Their visualization expertise contributed to many useful suggestions for improvements. In the last iteration, SAM was evaluated with data tracked over an entire semester. SAM allows the teachers to find students doing well and at risk, as well as getting a better course overview and understanding student time spent. Almost all participants over all studies would like to continue using SAM. We can conclude that SAM is useful to teachers and assists with improving awareness, based on the wide variety of new insights gained, the stated benefits and the eagerness to continue using SAM. The successful evaluations of SAM in four very diverse real-world environments illustrates the potential applicability in many learning settings.

Based on the interest and feedback we received, we believe that there is a large potential for visualizations of tracking information, not only to assist teachers and students, but also in other domains like personal Informatics [3] and Quantified Self⁹. Applications in these domains help people to understand their habits and behavior, through tracking and visualization, e.g. for sleep tracking and email communication patterns. Tracking of health data can improve for example fitness [20] and enable early diagnosis of illness. The rise of Big Data also requires visual analysis tools [31]. Within companies, awareness visualizations are used to analyze business and manufacturing processes [22]. Tools similar to SAM can

⁹Quantified Self, <http://quantifiedself.com/>

assist in all previously mentioned settings by visualizing the tracked data and extra metrics can be included, based on the context. Initially maybe some functionality can be perceived as somewhat complex, but motivated users will find the flexible mechanisms to view and manipulate trending data of great value. Validating SAM's impact in these fields requires further evaluations in these settings.

The methodology and results of this paper can be useful for other researchers when creating a visualization tool. Evaluating iteratively, early and in realistic settings has proven very valuable for us. In contrast to many existing works on the development and deployment of visualizations [19], the approach enabled us to gain insight into the usability and usefulness for real users. The methodology can be easily applied in other domains, e.g. personal informatics. Hopefully this work will inspire others to do more extensive user studies on visualizations. Not only usability studies of how people operate visualizations are needed, but also the analysis of how people really use those tools and how the tools impact their behavior are of great value. Providing different ways to visualize data can be useful for users. This allows users to select the visualization they feel comfortable with. Of course there is a trade-off: users will require more time to understand multiple visualizations versus only one.

A limitation of our studies is that we only exposed users to SAM for a brief time period. We will evaluate SAM over a longer period in the same Computer Architecture course this year and conduct multiple surveys during the course with both students and teachers. This will allow us to investigate the impact on students and teachers over time in detail. For this planned iteration, we will improve SAM with the most requested features and metrics, e.g. extend the stats window and add metrics of groups.

Acknowledgements

This research is funded by the European Commission's 7th Framework Programme (FP7/2007-2013) under grant no 231396 (ROLE). Katrien Verbert is a Postdoctoral Fellow of the Research Foundation - Flanders (FWO).

References

- [1] Distance learning: From multiple snapshots, a composite portrait. *Computers and Composition*, 23(1):139 – 161, 2006.
- [2] 1st International Conference on Learning Analytics and Knowledge 2011, February 2011.
- [3] Personal Informatics & HCI , CHI'11 workshop.
- [4] H. S. Alavi et al. Distributed awareness for class orchestration. *ECTEL '09*, pp 211–225, 2009.
- [5] K. R. Allendoerfer. *How information visualization systems change users' understandings of complex data*. PhD thesis, Drexel Uni., 2010.
- [6] A. Bangor, P. T. Kortum, and J. T. Miller. An empirical evaluation of the system usability scale. *Int. Journ. of HCI*, 24(6):574–594, 2008.
- [7] J. Benedek and T. Miner. Measuring desirability: New methods for evaluating desirability in a usability lab setting. In *Proc. of UPA Conf.*, 2002.
- [8] J. Brooke. SUS: A quick and dirty usability scale. In *Usability Evaluation in Industry*. 1996.
- [9] A. Dasgupta and R. Kosara. Pargnostics: Screen-space metrics for parallel coordinates. *IEEE TVCG*, 16:1017–1026, 2010.
- [10] S. Govaerts et al. Evaluating the student activity meter: two case studies. In *ICWL, LNCS*, in press, 2011.
- [11] S. Govaerts et al. Visualizing activities for self-reflection and awareness. In *ICWL, LNCS*, 6483, pp 91–100, 2010.
- [12] J. Hardy, S. Bates, J. Hill, and M. Antonioletti.

- Tracking and visualisation of student use of online learning materials in a large undergraduate course. In *ICWL, LNCS*, 4823, pp 464–474, 2008.
- [13] J. Janssen et al. Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Computers & Education*, 49(4):1037–1065, 2007.
- [14] D. Keim et al. Visual analytics: Definition, process, and challenges. In *Information Visualization, LNCS*, 4950:154–175. 2008.
- [15] T. Z. Keith. Time spent on homework and high school grades: A large-sample path analysis. *Journal of Edu. Psycho.*, 74(2):248–253, 1982.
- [16] R. Mazza and C. Milani. Exploring usage analysis in learning systems : Gaining insights from visualisations. *Time*, p 65, 2005.
- [17] J. Nielsen. *Usability Engineering*. 1st ed., 1993.
- [18] Ž. Obrenović. Design-based research: what we learn when we engage in design of interactive systems. *Interactions*, 18:56–59, 2011.
- [19] C. Plaisant. The challenge of information visualization evaluation. *AVI '04*, pp 109–116, 2004.
- [20] S. Purpura et al. Fit4life: the design of a persuasive technology promoting healthy behavior and ideal weight. In *CHI '11*, pp 423–432, 2011.
- [21] H.-C. Schmitz et al. Camera for PLE. In *EC-TEL, LNCS*, 5794, pp 507–520. 2009.
- [22] M. Sedlmair et al. Cardiogram: visual analytics for automotive engineers. In *CHI '11*, pp 1727–1736. 2011.
- [23] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *IEEE VL/HCC*, p 336. 1996.
- [24] H. Siirtola et al. Visual perception of parallel coordinate visualizations. In *IV '09*, pp 3–9. 2009.
- [25] J. Sinclair et al. The folksonomy tag cloud: when is it useful? *J. of Info. Science*, 34:15–29, 2008.
- [26] A. Soller et al. From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. *I. J. AI in Education*, 15(4):261–290, 2005.
- [27] D. Straker. *A Toolbook for Quality Improvement and Problem Solving*. Prentice Hall, 1995.
- [28] T. S. Tullis et al. A comparison of questionnaires for assessing website usability. In *UPU Conf.*, 2004.
- [29] M. van Someren et al. *The think aloud method: a practical guide to modelling cognitive processes*. 1994.
- [30] J. Wang et al. Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets. In *IEEE INFOVIS*, pp 105–112, 2003.
- [31] W. Wang et al. Visualization of large hierarchical data by circle packing. *CHI '06*, pp 517–520, 2006.
- [32] M. Q. Wang Baldonado et al. Guidelines for using multiple views in information visualization. In *AVI '00*, pp 110–119, 2000.
- [33] M. Wolpers et al. Tracking actual usage: the attention metadata approach. *Educational Technology & Society*, 10(3):106–121, 2007.
- [34] H. Zhang et al. Moodog: Tracking students' online learning activities. In *EDMEDIA '07*, pp 4415–4422.
- [35] J. Zhao et al. Kronominer: using multi-foci navigation for the visual exploration of time-series data. In *CHI '11*, pp 1737–1746, 2011.
- [36] H. Zhou et al. Splatting the lines in parallel coordinates. *CG Forum*, 28(3):759–766, 2009.
- [37] B. J. Zimmerman. Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2):65–70, 2002.