Inferring Higher Level Learning Information from Low Level Data for the Khan Academy Platform

Pedro J. Muñoz-Merino, José A. Ruipérez Valiente, Carlos Delgado Kloos
Universidad Carlos III de Madrid
Avenida Universidad 30, 28911 Leganés (Madrid) Spain
{pedmume, jruipere, cdk}@it.uc3m.es

ABSTRACT
To process low level educational data in the form of user events and interactions and convert them into information about the learning process that is both meaningful and interesting presents a challenge. In this paper, we propose a set of high level learning parameters relating to total use, efficient use, activity time distribution, gamification habits, or exercise-making habits, and provide the measures to calculate them as a result of processing low level data. We apply these parameters and measures in a real physics course with more than 100 students using the Khan Academy platform at Universidad Carlos III de Madrid. We show how these parameters can be meaningful and useful for the learning process based on the results from this experience.

Categories and Subject Descriptors
K.3.1 [**Computer Uses in Education**]: *Distance learning;* H.1.2 [**User/Machine Systems**] *Human information processing*

General Terms
Algorithms, Measurement, Experimentation, Human Factors

Keywords
Learning analytics, visualization, hints

1. INTRODUCTION
An analogy can be established between learning and the business sector, where the use of data mining techniques and business intelligence tools [1] has become widespread in companies. A big issue for business intelligence is how to deal with thousands of data that are difficult to understand and convert these into high level meaningful information that can be used as a basis for decision making by organizational stakeholders [2]. We are looking for the same outcome in the e-learning area.

We understand low level learning data to be the collection of event entities and all their related data (e.g., time or resources involved), which are usually stored in a database. They do not usually convey any meaningful sense alone, but if we process them properly, then useful information can be obtained. Our motivation for this work is to transform a huge amount of low level learning data into high level parameters that can be meaningful for students and teachers, in order to answer questions such as: Can this user be motivated by gamification techniques?

LAK '13, April 08 - 12 2013, Leuven, Belgium.

In this work, we define a collection of high level learning parameters that give insights into the learning process. These parameters are calculated based on low level data. We have applied these parameters to the Khan Academy[1] platform, extending the Khan Academy learning analytics module. Some cases and results show the importance and meaningfulness of these parameters for the learning process. We illustrate them in our case study on a physics course using the Khan Academy platform with more than 100 students.

2. RELATED WORK
The collection of low level educational data is very important. Approaches such as Contextualized Attention Metadata (CAM) [3] allow the retrieval of all the events from distributed sources. The data can be collected in different formats, such as the Resource Description Framework (RDF) [4].

Low level educational data as well as high level information have been reported in different works. Some of the approaches [5, 6] focus on the prediction of events by applying data mining techniques and statistical methodologies. Other works present practical specific visualization tools such as goals, activities, or number of events per item [7], social network visualizations in the CAMERA tool [8], or resources used, average time spent per resource, or the evolution of the students in the SAM tool [9]. However, there are many new high level information parameters that have not been addressed by the literature.

We present some high level information parameters that have been applied to the Khan Academy platform. These parameters were not present in the Khan Academy learning analytics module, so we extended it. Some of the parameters relating to exercise solving habits (e.g., hint abuse) have already been presented in the literature (e.g. in [10]), but we have adapted the method of calculating them as the semantic of the Khan Academy platform is different from the Geometry Cognitive Tutor framework [10].

3. EDUCATIONAL ENVIRONMENT
An instance of the Khan Academy software is installed on our own application server[2]. This instance is personalized with our own pages, badges, etc. Specific materials, including videos and exercises, were developed at Universidad Carlos III de Madrid for a course on physics and were uploaded onto our instance of the Khan Academy software. The course comprised a complete set of 27 videos and 35 exercises. This course took place in August 2012 over the entire month. More than 100 students were registered on the course, which is a review of prerequisites that the students should know before starting an engineering degree.

[1] https://www.khanacademy.org/

[2] http://uc3m-ka.appspot.com/

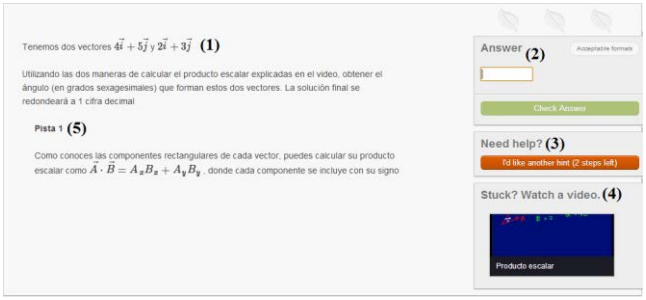

Figure 1: Exercise interface at Khan Academy.

Figure 1 shows an exercise example based on the scalar product running on our own Khan Academy instance. Each exercise has a related video (4) and usually some hints (3). Exercises can be parametric (1). When a student accesses an exercise, this exercise will not change (even if it is accessed at different moments) until the student solves it correctly. The next time that the exercise is accessed, a new exercise on that same topic is obtained, with only the parametric values changing. In order to be 'proficient' at a certain exercise, considering maximum efficiency, a student must complete eight exercises correctly in a row without asking for any hint (5) and answering correctly at the first attempt (2).

Khan Academy also incorporates gamification aspects such as points that users can earn. Moreover, a student can win badges, e.g., by achieving proficiency in a certain group of exercises or topics.

In order to calculate these high level parameters, a set of python scripts using the Google App Engine (GAE) API were developed since Khan Academy runs over GAE infrastructure.

4. PROPOSED PARAMETERS

In this section we propose a set of high level information parameters and show how to obtain them from the processing of the low level data. The parameters are grouped into five blocks.

4.1 Total Use of the Platform

These parameters do not take into account if a user has done very well or badly, but only the total use of the platform.

4.1.1 Total Effective Use of the Platform (TUP)

We denote the total available exercises and videos as AE and AV, respectively. In the case of videos, we use two measures: one for completed videos (CV) and another for started videos (SV):

$$TUP_1 = \frac{\text{CV}}{\text{AV}} \ ; \ TUP_2 = \frac{\text{SV}}{\text{AV}}$$

Three measures are proposed for exercises relating to: the number of different types of exercises a student attempted (DEA), that a student spent enough time (DET), and that a student attempted at least some number of times or solved correctly (DEN).

$$TUP_3 = \frac{\text{DEA}}{\text{AE}} \ ; \ TUP_4 = \frac{\text{DET}}{\text{AE}} \ ; \ TUP_5 = \frac{\text{DEN}}{\text{AE}}$$

Moreover, a global measure of total use of videos plus exercises is possible, taking into account a weight for different videos and exercises for all presented measures.

4.1.2 Total Efficient Use of the Platform (TEP)

In the case of videos, a measure is proposed: dividing the maximum possible video time length (TVL) by the total time spent on completed videos by the user (TEV). Therefore, if a user repeats watching a video, his efficiency will decrease:

$$TEP_1 = \frac{\text{TVL}}{\text{TEV}}$$

Two parameters are proposed for exercises: the total number of exercises attempted (TEA) divided by DEA gives a measure of the number of times a user repeats his exercises on average; the total time invested in all exercises (TTE) divided by the total normal estimated time (TNT) gives a measure of whether the student spent more time than expected, on average, on solving exercises.

$$TEP_2 = \frac{TEA}{\text{DEA}} \ ; \ TEP_3 = \frac{TTE}{\text{TNT}}$$

4.1.3 Total Time and Use of Optional Items

The total time on platform (TT) is the time (in minutes) that a user has spent with exercises (TTE) and videos (TTV). This time can exceed the real time a user has spent interacting with videos and exercises, because they can have the video or exercise windows open without watching videos or doing exercises.

Moreover, the platform offers several features that are not mandatory to use in an educational environment, e.g., for our Khan Academy environment the profile personalization or the establishment of goals. We measure with this parameter if the student is motivated by features of the platform other than the normal ones. A simple example of this measure can be to distinguish people who had some interaction with an optional item from people who did not.

4.2 Correct Progress on the Platform

This block contains parameters that represent how well users have interacted with the platform. This block does not take into account the total use but the performance of the student with the materials.

4.2.1 Effective Correct Progress on the Platform (ECP)

Correct progress on the platform for videos can be given as the percentage time of all the videos that have been watched, although other video measures presented for total use might also be used.

For exercises, three measures are proposed: ECP_1 is the total correct exercises (TCE) divided by the minimum total number of correct exercises to obtain proficiency (8 in the case of the Khan Academy platform) and multiplied by AE, with a limit so that TCE for each type of exercise cannot be greater than 8; ECP_2 is the number of proficiencies achieved (UP) divided by AE; ECP_3 is the total progress, that is the average of the progress on all exercises, to obtain proficiency. Each exercise has a related progress for each student from 0 to 1 depending on the number of correct exercises, incorrect attempts or hints used in that type.

$$ECP_1 = \frac{TCE}{8 * AE} \ ; ECP_2 = \frac{\text{UP}}{\text{AE}} \ ; \ ECP_3 = total\ progress$$

4.2.2 Efficient Correct Progress on the Platform (EP)

The efficient correct progress on the platform for videos can be the same as for the efficient total use. Regarding exercises, one measure is defined as the division of the total different types of exercises correctly solved (TDCE) by the total time spent in solving them (TTE) and multiplied by the average expected time to solve an exercise. This time has been set to 40 seconds for all the exercises of our Khan Academy educational environment. Another measure is the division of TCE by TEA:

$$EP_1 = \frac{TDCE * \ 40\ seconds}{TTE}; \ EP_2 = \frac{TCE}{TEA}$$

In addition, a maximum time limit is set for the contribution of each exercise to TTE (180 seconds for our case) so that noise is not introduced for the measure. EP_1 values considerably less than 1 indicate non-efficient users, and values considerably greater than 1 indicate that users solve many exercises correctly in a reduced amount of time.

4.3 Time Distribution of Use of the Platform

This section represents an analysis of the times when the user interacted with the platform.

4.3.1 Total Working Schedules

These parameters show at what time users watch their videos and do exercises. We set three time schedules (TS) by time intervals: morning [7:00 to 13:59] (TM), afternoon [14:00 to 20:59] (TA) and night [21:00 to 06:59] (TN). The percentages of use in each time schedule are calculated.

4.3.2 Efficiency by Working Schedules

These parameters use the same time schedule as in the last subsection, but in this case the efficiency (ESP) of the user doing exercises is measured in each time interval.

$$ESP_1 = \frac{TCE\ morning}{TEA\ morning}\ ;\ ESP_2 = \frac{TCE\ afternoon}{TEA\ afternoon};$$

$$ESP_3 = \frac{TCE\ night}{TEA\ night}$$

We propose this measure aiming to find that some users might work better at different times of the day.

4.3.3 Constancy of Users

This parameter checks if a user was studying in a constant way during several days or was studying strongly only for a few days. The sample mean and the variance of time spent on the platform by day are calculated:

$$\overline{m}_{time/day} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad ; \quad \sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \overline{m})^2}{(N-1)}$$

N being the number of days that the course took and *x* the time spent each day. The learning constancy is calculated using the sample variance of the time from each day, i.

4.4 Gamification Habits

Here we try to analyze whether a user is motivated by the gamification elements. A measure related to user badges (UB) is proposed. This parameter consists of the total number of badges that the user has earned (EB) divided by the exercise correct progress on the platform (ECP). A user that achieves more badges than another on the platform (their correct progress on the platform being the same) will be more motivated because of the badges. In addition, if two users have earned the same number of badges but have different exercise correct progress on the platform, the one with better correct progress on the platform will be less motivated by badges, because the more a user advances on the platform correctly, the more badges he can earn.

$$UB = \frac{EB}{ECP}$$

4.5 Exercise Solving Habits

These parameters represent the way a user solves an exercise.

4.5.1 Explorer or Recommendation Listener

Khan Academy allows users to define prerequisites between exercises and also includes an exercise recommender. Therefore, checking whether a user has accessed a certain number of exercises according to the system recommendation will give us an indication of whether a user usually follows the recommended learning path (RL):

$$RL = \frac{recommended\ TEA}{TEA}$$

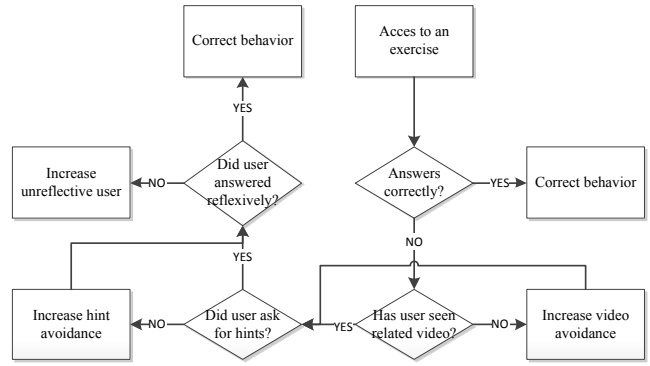

Figure 2. Model flow diagram.

4.5.2 Hint Avoidance, Video Avoidance, Unreflective User and Hint Abuser

We propose a model that tries to cover all the possible situations that a learner can encounter when interacting with an exercise on the Khan Academy platform. Figure 2 shows this model. The flow diagram starts at the point where a user is answering an exercise. If the user answers correctly, then the counter of correct behavior is increased. If the answer is wrong, then the system checks whether the student had watched the related video. If he did not watch it, then the local value for video avoidance (VA) profile is set to 1, otherwise is set to 0. Next, the system checks whether the student requested for hints. If he did not, the local value for hint avoidance (HAV) profile is set to 1, otherwise it will be a number in the range [0,1] representing the percentage of requested hints for that type of exercise. Lastly, if the student answers too fast, (e.g., less than 10 seconds), then the local value for unreflective user (US) profile is set to 1, otherwise is 0. It is important to note that we apply this profile only until the student does one exercise of that type correctly; otherwise we would be contaminating with noise since the user already knows how to solve the problem and subsequent similar problems as they are parametric.

Each time that a student accesses a different type of exercise, each of the aforementioned parameters is set from [0, 1]. This is the local value of the parameters for a type of exercises. The global value of these parameters is the mean of all the local ones among the exercises in which students had some interaction and where these parameters were set to some value.

In addition, the hint abuser parameter takes into account the exact time for the first hint as well as time intervals between hint requests. For example, if a user starts an exercise and in less than 10 seconds he has already requested a hint, then the hint abuser counter is increased. Similarly, if the time difference between two hint requests is lower than 10 seconds the hint abuser counter is also increased.

5. RESULTS AND DISCUSSION OF THE MEANING OF PARAMETERS

In this section, we show how the presented parameters have a utility in the learning context. Moreover, we present typical situations in our course where they can be used. We analyze their meaning in the context of our Khan Academy educational environment and present some illustrative results.

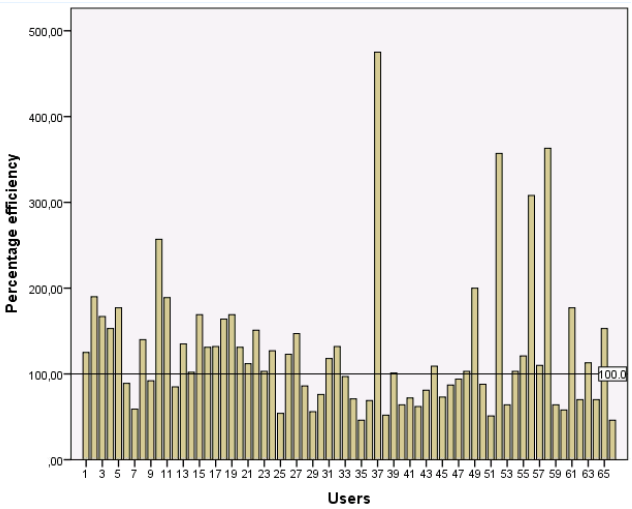

Figure 3. Exercises efficiency in time.

Although a total of more than 100 students interacted in the course, we have not taken into account the ones who we consider that did not interact enough; therefore the following results are based on the analysis of 66 students.

The plan for the physics course is based on the methodology of flipping the classroom, so that students prepare the lessons using the Khan Academy platform in advance of face to face lessons. It is important for teachers to know which students are prepared well for the face to face lessons; therefore the parameters for the effective correct progress on the platform are important. Based on these parameters, teachers can set a threshold for considering which students are prepared well, e.g., a minimum number of completed videos or minimum number of exercises where proficiency is obtained. Usually, teachers will set a combination of conditions on these parameters to consider that a student is well prepared for the face to face course (e.g., an AND of conditions or a global measure taking everything into account with a weight).

Considering a minimum of 16 videos totally completed (from TUP_1) or 21 exercises where a student obtained proficiency (from ECP_2), we can say that 22 out of 66 students did well on the platform and were well prepared for the face to face sessions.

Among the students who did well on the platform (passed this threshold of correct progress), the measures of effectiveness give us an idea of who was more effective in terms of time or less repetition of items. Students who are not efficient in their learning might be advised or guided so that they can take better advantage of their time, because it is not only important to learn but to do so in an efficient way.

The parameters relating to effectiveness can be given in the form of graphs. Figure 3 shows a representation of EP_1 for the exercise solving time efficiency for all students in our experiment. The middle value line set upon 100 percent would be the average time that a normal user should spend to solve the exercises. If students are far above that line, it means that they solve the exercises faster than the critical value, while students below the line need more time per correct exercise. Therefore, two students can have similar values of correct progress but one can be more efficient in time (or number of attempts) than the other.

Among the students that did wrong on the platform (did not pass the threshold of correct progress), the parameters of total use of the platform will let us know whether the students made some effort to learn and where (videos, exercises, time) or if they did not make any effort. For example, for a time (TT) exceeding 225 minutes of interaction with the platform, and more than 15 started videos (from TUP_2), or more than 20 attempts at different types of exercises (from TUP_3), we can detect that 8 out of the 44 students that did badly made a considerable effort on the platform. These students might need more remedial support.

A Pearson correlation test shows that there is a statistically significant difference at 99% level between the total time (TT) and the following parameters: videos completed from TUP_1 (r=0.80), videos started from TUP_2 (r=0.81), exercises attempted from TUP_3 (r=0.71), and exercises with proficiency from ECP_2 (r=0.73). Therefore, the total time is related strongly to these measures, and it is a good parameter to predict the number and quality of interactions with the platform.

Another important issue is to identify whether or not students are motivated to achieve badges, and this is indicated by the gamification habits parameters. Two students might have a strong activity on the platform, but one of them might have a lot of badges while the other has only a few badges, indicating that he is not motivated by them. Students that are motivated by gamification can be identified and participate in future gamification activities.

The parameter of total use of optional elements gives information about whether students were interested in extra functionalities of the platform that were not mandatory and about which they were not given any information. A total of 17 students used some type of optional functionality. This may denote curiosity and identify students who like to explore things. The Pearson correlation between the use of optional items and the total time (r=0.16, p=0.19) and the percentage of proficiencies obtained (r=0.3, p=0.014) suggest that the use of optional items, or not, is not strongly related to the total time of use of the platform or whether or not the user obtains proficiency in exercises.

In addition, surprisingly, there was not a statistically significant relationship between the use, or not, of optional items and the recommender/explorer parameter from RL (r=0.1, p=0.42). One might tend to think that students who use optional items would tend to be explorers and not take into account the recommendations, but this relationship was not found in the experiment.

The parameters relating to constancy in learning give us an indication of whether or not students learn in a constant way. The variance—but also the mean—should be taken into account for the interpretation. In many situations, students might learn better for the long term if they do it in a constant way, so a system might recommend non-constant students to learn in a more constant way or hide some activities from them until some specific date.

Figure 4 shows the constancy measures (mean and variance) applied to top users according to their activity on the platform during a time interval from [01/08/2012] to [08/09/2012]. We can see, for example, that user 4 is a constant student because his variance is very low with respect to the mean. In the same way, user 8 has not been learning in a constant way but only for a few days. With a similar analysis we can reach conclusions for all the users.

Moreover, time schedules where students spent more time and where they were more efficient can be of interest, e.g., for personalization of tasks to time slots.

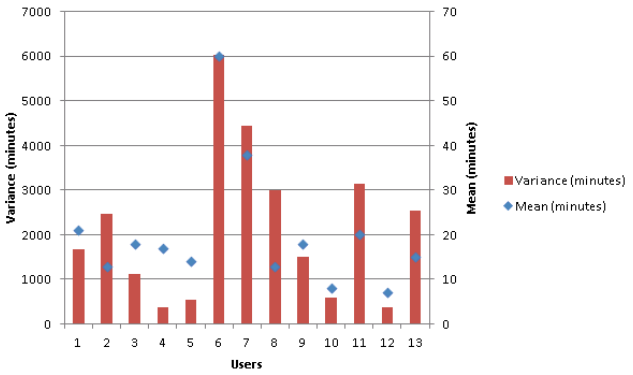

Figure 4. Constancy measure applied to top users.

Finally, some help-seeking bugs, such as help abuse, are correlated with learning gains according to the literature [10]. Therefore, knowing the exercise solving habits parameters is of special importance as teachers can act and intervene to change the behavior of students so that they can learn in a more effective way.

Considering a level of more or equal than 25% as the threshold for each of the problem solving habits parameters, we can say that 30.3 % of students had the profile of hint avoider, 25.8 % of video avoider, 40.9 % of unreflective user and 12.1% of hint abuser.

In addition, table 1 shows the correlations among the different categories of problem solving. The only correlations that are statistically significant at the 99% level are 1) the unreflective user with respect to hint abuser and hint avoider, which makes sense as a user that does not reflect on his learning can select, or not, a hint by chance, and 2) video and hint avoidance, which also makes sense as a user who avoids hints can also have tendency to avoid videos. In addition, the relationships that were not statistically significant also make sense, e.g., the hint abuser with respect to hint and video avoidance, as a user who tends to use abuser techniques will not tend toward avoidance ones.

Table 1: Correlations among problem solving habits

	Hint avoid.	Video avoid.	Unrefl. user	Hint abuser
Hint avoidance	1	0.382	0.607	-0.186
Video avoid.	0.382	1	0.289	0.096
Unrefl. user	0.607	0.289	1	0.317

6. CONCLUSIONS

This paper proposes a set of high level parameters that can give useful information for students and teachers about the learning process, and we have illustrated it with a case study of more than 100 students using the Khan Academy platform. The information obtained using our proposed parameters is not easy to obtain with the present Khan Academy learning analytics module. For example, if a teacher wants to know about problem solving habits, he must go through different windows to see which videos a student has watched, the time spent, and the details of the resolution of each exercise and make many complex calculations.

The proposed parameters can be applied not only for the Khan Academy platform but for other systems, incorporating the proper adaptations. In this direction, the semantics of each platform influences the types of parameters that can be used and the way to measure them, e.g., the correct use of exercises can be redefined in other platforms where the same exercise is not repeated to obtain proficiency. In addition, on some occasions some parameters cannot give useful information, for example, a student may obtain all the badges and solve everything correctly, but we cannot say that he is or is not motivated by the badges.

The results of this study can be applied for useful interventions, for example, in recommender applications. Another challenge is how to visualize all the information in an easy way for teachers.

7. ACKNOWLEDGMENTS

Work partially funded by the EEE project, "Plan Nacional de I+D+I TIN2011-28308-C03-01" and the "Emadrid: Investigación y desarrollo de tecnologías para el e-learning en la Comunidad de Madrid" project (S2009/TIC-1650)".

We would like to thank the professors in Universidad Carlos III de Madrid who developed the physics materials for the course (videos and exercises).

8. REFERENCES

[1] P. Baepler and C. J. Murdoch. Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning*, 4(2), July 2010.

[2] H. Drachsler, and W. Greller. The pulse of learning analytics understandings and expectations from the stakeholders. In *LAK '12 Conference Proceedings*, pages 120–129, 2012.

[3] M. Wolpers, J. Najjar, K. Verbert, and E. Duval. Tracking actual usage: the attention metadata approach. *Educ. Technol. Soc.*, 10(3):106–121, 2007.

[4] P. J. Muñoz-Merino, M. Wolpers, K. Niemann, M. Friedrich, A. Pardo, C. Delgado Kloos, and M. Muñoz-Organero. CAM in the semantic web world. In *I-SEMANTICS '10 Conference Proceedings*, 2010.

[5] L. P. Macfadyen, and S. Dawson. Mining LMS data to develop an ''early warning system" for educators: A proof of concept. *Computers & Education*, 54(2):588–599, February 2010.

[6] C. Romero, S. Ventura, and E. García. Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1):368–384, August 2008.

[7] E. Duval, J. L. Santos, K. Verbert, and S. Govaerts. Goal-oriented visualizations of activity tracking: a case study with engineering students. In *LAK '12 Conference Proceedings*, pages 143–152, 2012.

[8] H. Schmitz, M. Schefel, M. Friedrich, M. Jahn, K. Niemann, and M. Wolpers. CAMera for PLE. In *EC-TEL '09 Conference Proceedings*, pages 507–520, 2009.

[9] S. Govaerts, K. Verbert, E. Duval, and A. Pardo, The student activity meter for awareness and self-reflection. In *CHI EA '12 Conference Proceedings*, pages 869–884, 2012.

[10] V. Aleven, B. M. McLaren, O. Roll, and K. Koedinger. Toward tutoring help seeking; Applying cognitive modeling to meta-cognitive skills. In *ITS-2004 Conference Proceedings*, pages 227-239, 2004.