Journal of Universal Computer Science, vol. 21, no. 1 (2015), 134-155
submitted: 16/5/14, accepted: 1/10/14, appeared: 1/1/15 © J.UCS

A Visual Analytics Method for Score Estimation in Learning Courses

Luis de-la-Fuente-Valentín
(Universidad Internacional de La Rioja, Logroño, Spain
luis.delafuente@unir.net)

Abelardo Pardo
(The University of Sydney, Australia
abelardo.pardo@sydney.edu.au)

Fernando López Hernández
(Universidad Internacional de La Rioja, Logroño, Spain
fernando.lopez@unir.net)

Daniel Burgos
(Universidad Internacional de La Rioja, Logroño, Spain
daniel.burgos@unir.net)

Abstract: The provision of awareness is a well-known method for fostering students' self-reflection, a metacognitive skill often related to academic success and considered one of the key skills of the 21st century. Although the information discovered using learning analytics techniques can be useful in fostering self-reflection, its delivery to students should be done without distracting them from their learning goals. This paper presents a visualization technique based on similarity measures and their relationship with final course results, in order to foster students' awareness. The approach is based on the idea that 'students that behave similarly are graded similarly'. This idea is validated with an empirical evaluation to determine the visualization technique's accuracy when used to find a relationship between similarity and grade. The study used a previously collected dataset and several volunteers were asked to estimate the students' scores with graphics provided as the only source of information. The obtained results validate the proposal as a means to foster effective self-reflection.

Keywords: human-computer interface; intelligent tutoring systems; learning analytics; self-reflection; score estimation; awareness
Category: L.3, L.2.4, L.2.1

1 Introduction

Self-reflection, understood as the ability to evaluate the results of one's own learning efforts, is considered to be an important factor in effective learning as it promotes the construction of more complex, related and integrated knowledge structures [McCrindle & Christensen, 1995]. The ability to self-reflect and self-regulate is related to success [Ertmer & Newby, 1996] and it is also considered one of the so-called '21st century skills' [Binkley et al., 2010]. However, metacognitive skills can be difficult to achieve [Lew & Schmidt, 2011], and as a consequence there are numerous research initiatives studying how to promote successful self-reflection. For example, the work presented by [Wang et al., 2011] uses knowledge visualization techniques to support resource-abundance for self-regulated learners. Many other systems also use visual approaches as the means to reach self-reflection. [Santos et al., 2013] presents a system that visualizes different learning logs and evaluates their impact on students' habits. Another example is given by [Govaerts et al., 2012], which presents visualization methods to analyse trending data. In those cases, the information obtained from the course is pre-processed and visually presented to the students in order to allow them to self-assess their knowledge and/or performance.

Awareness - defined as the provision of knowledge about the self, the group and one's context - pushes learners to meditate on their circumstances, that is, to self-reflect. Pushing the learner to self-reflect is precisely the goal of [Carmean & Mizzi, 2010], whose proposal is to "influence learners' actions without infringing on their freedom of choice." A similar idea uses analytics to provide users with what they need without being explicitly asked to do so [Romero & Ventura, 2011]. In practice, this translates into self-reflection being frequently enabled by means of awareness mechanisms. For example, by evaluating how visual awareness modifies attitudes towards engagement in online communities [Glahn et al., 2009], or by analysing how people perceive different visualization methods, we can use knowledge awareness maps to support learners in a ubiquitous learning experience [El-Bishouty et al., 2007]. Given that the field of visualization methods is so relevant for supporting teachers and learners, [Santos et al., 2013] and [Verpoorten et al., 2011] discuss the concept of 'learning dashboards', and identify its empirical evaluation as a research challenge. Also, [Leony et al., 2012] present a framework for the creation of learning dashboards, considering different roles in the learning process and allowing for their integration with third party systems.

Awareness provision methods have recently been advanced by the upcoming research in the area of learning analytics. According to the International Conference on Learning Analytics and Knowledge, learning analytics refers to "the measurement, collection, analysis and reporting of data about learners and their contexts, for [the] purposes of understanding and optimizing learning and the environments in which it occurs" [Siemens, 2012], and it is one of the fastest growing fields in technology enhanced learning research [Baker & Yacef, 2009]; [Ferguson, 2012]; [C. Romero & Ventura, 2007]. Due to the ability of learning management systems (LMSs) to track student actions, distance online learning is a well-suited area of application for learning analytics. [McGrath, 2010] and [Romero et al., 2008] provide sample use cases of how user activity tracking challenges are met with data mining techniques in the context of LMSs.

When analytics techniques make use of information visualization techniques, this is typically referred to as 'visual analytics'. More formally, visual analytics techniques use graphical representations to synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data to detect the expected and discover the unexpected [Keim et al., 2008]. The idea behind visual analytics is to let a computer program filter and pre-process the data, arrange it visually and then let the user perform an interpretation. For instance, the work presented by [Larusson & Alterman, 2009] and [Mazza & Dimitrova, 2004] support teachers in keeping track of students' participation with graphical representations of their activity, while [Donovan et al., 2008] use visual analytics to obtain visual recommendations.

Analytics techniques are also used for the early identification of students at risk and score prediction, as a straightforward way to help students self-assess their performance in a course. This problem is addressed by a field known as 'academic analytics' [Campbell et al., 2007], with many examples in the literature: [Macfadyen & Dawson, 2010] presents an "early warning system" for educators that mines data from the LMS; [Munoz-Organero et al., 2010] establishes a relationship among LMSs' usage patterns and students' motivation; and [Romero-Zaldivar et al., 2012] analyse the correlation between students' involvement on a course and the score they obtain.

However, due to the potentially high impact of incorrect predictions, existing techniques are rarely targeted at students and instead are teacher-oriented [Cocea & Weibelzahl, 2009]. In a recent review of existing Learning Dashboards, [Verbert et al. 2013] identifies that only 4 from the 15 reviewed dashboards are targeted at the student. Furthermore, awareness provision mechanisms are usually based on the presentation of tracked data, with almost no presence of data mining mechanisms to infer hidden information from the tracked data. Another drawback of current systems is the required input data: they frequently make use of demographic variables and academic records from previous courses [Potgieter et al., 2010], but this data is only available under very restricted circumstances. In order for these prediction schemes to gain widespread adoption, there is a need for prediction schemes that rely solely on data collected locally during one course.

This paper proposes a visualization method based on the similarity between behavioural patterns and their relationship with course results. Students are presented with a graphic that identifies similar students from previous course editions and reports their score. The students can therefore self-assess their progress by comparing their performance with others. As presented in [Verbert et al. 2013], most of the learning dashboard are limited to the presentation of raw tracked data. Therefore, the use of data mining techniques to identify similar students and the use of visual techniques to establish a relashionship between similarity and obtained score is, to the best knowledge of the authors, a step forward in the state of the art. The idea is based on the concept of the 'nudge' [Carmean & Mizzi, 2010] and its goal is to increase students' self-awareness by helping them self-assess their progress in the middle of a course and avoid any misalignment of expectations at the end of the course. Other similar works are presented are CourseVis [Mazza & Dimitrova, 2007], SAM [Govaerts et al. 2012], Step Up [Santos et al. 2013] and GLASS [Leony et al., 2012]. Among them, the advantage of the presented approach is the identification of similar students from previous courses as a method to self-estimate the students' performance. For the empirical validation of the proposed visualization technique, several volunteers were asked to estimate the students' scores with the graphics provided as the only source of information. The goal was to determine the accuracy of the technique by categorical and numerical score estimation, and to measure its perceived usefulness and perceived ease of use [Davis, 1989].

The rest of this paper is organized as follows: Sections 2 and 3 detail the proposed visualization technique, including analytics techniques and representation methods. Section 4 explains the methodology for the empirical validation of the proposal. The experimental results are presented and discussed in Section 5 and, finally, Section 6 draws some conclusions and poses research questions to be addressed in future work.

2 The Proposed Visualization Technique

This section details an innovative technique of visual analytics for use in educational scenarios.

2.1 Visualization Description

The proposed visualization technique is based in the concept of ‘similarity’ between student behaviours. The reason for this approach is that, in our experience and observations, students tend to have a better notion of their performance when comparing it to somebody else taking part in the same course, or in previous editions of the same course. A visualization enables the students to more easily interpret the information, rather than being given a single numerical value summarizing the assessments carried out so far, a method which fails to convey the idea of a trajectory and a position within a cohort of students.

The idea can be illustrated with the following example: if a teacher says that “this exercise is not mandatory, but the statistic says that those students who completed it in previous courses got the best overall scores in the course”, it can be foreseen that the students that want the highest scores will try to complete the exercise. That is, they will try to be “similar” to the best students in previous courses. With the same idea, the presented visualization identifies similar students from previous courses and reports the results they achieved.

The proposal uses scatter plots, with each point representing a different student. Such a two-dimensional approach allows for the easy visual recognition of clusters, without overloading the user with exhaustive information. The scheme assumes that students participate in two types of assessed activities (practical, and theoretical, in the presented case), allowing for the identification of the horizontal and vertical axes with these two activity-types, and thus making it easy to calculate a given student’s score. A different classification of activities is also possible. In the presented case, a student’s position in the graphic is given as follows:

- The horizontal axis represents the student's score during practical activities (e.g., homework, programming exercises and/or hands-on sessions).
- The vertical axis represents the student's score for theoretical activities (e.g., mid-term or final exams).

The total score for a student is represented by a position in the graph, and calculated by adding the value of the horizontal and vertical coordinates.

The similarity to other students is represented with different colour intensities, and the actual value depends on the student used for the comparison (e.g., student *A* can be similar to student *B*, but dissimilar to student *C*). In other words, the graphical representation of such similarity will be different for each student in the course. In practice, all the points in the graphic are the same colour (green, in the example), but with different intensities. When the graphic is presented to a student, she is taken as the reference and the colour intensities of the rest of the points are correspondingly applied: the darker the colour, the more similar the student. When presented to a different student, the similarity measures will offer different values and, therefore, the graphic will be repainted with the points in the same position, but with different intensities. That is, for the purpose of this paper, we use the “reference student” as “the student the graph is computed for”. The similarity measurement method is described in Section 3.

It is important to note that the colour intensities are obtained from the behavioural similarity measures, while the student's position in the graph is given by her own score. Strictly speaking, these are two independent variables and there is no need to find a relationship between colour intensities and position. However, it is foreseeable that students with similar behaviours might also obtain similar scores, and this suggests that similar students should be located nearby in the graph. The empirical evaluation of this assertion is presented in Section 5.

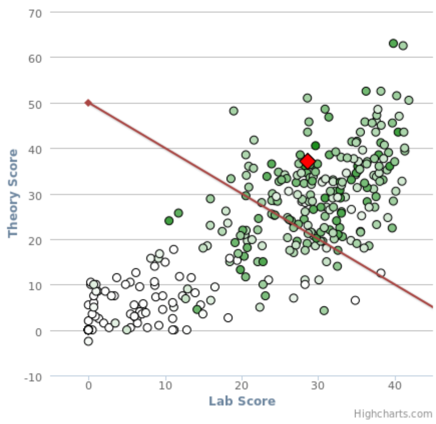


Figure 1: Visualization including the reference student

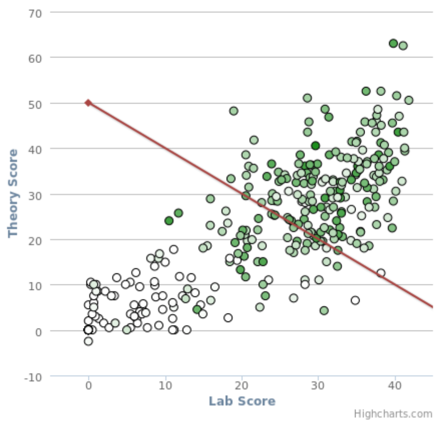


Figure 2: Proposed visualization technique to make score estimations

Figure 1 used a dataset from an already finished course, so the location (i.e. grades) of the reference student was known. In Figure 1, the reference student is represented as a red diamond and the rest of the students are green circles, ranging from dark-green to light-green/white. The red line is the pass/fail threshold for the course: students above the line passed the course. Figure 2 represents the same example, but without drawing the reference student. The red diamond is only possible to show because the dataset corresponds to an already finished course, so we know her grade (we are extracting this student from the dataset using him as a student in the current course). Therefore, Figure 1 is useful only to explain how the visualization technique works, but the red diamond will not be shown to the students in the real settings. The actual proposed technique is represented by Figure 2, which allows for making a rough estimation of ‘where the red diamond could be’, and the presented empirical validation is based on it. Section 2.3 details how the visualization technique can be used to make score estimations.

2.2 Considering Information in Chronological Order

The above-proposed visualization technique creates graphics in which similarity is calculated from the total number of events in a certain period of time, and so discards the chronological order of the events; however, this information might be relevant to more accurately determine whether two students have similar behaviour. In order to show how time affects the visualization method, the dataset was split into temporal sub-datasets, corresponding to different periods of the course. Figure 3 shows how the similarity values differ from one sub-graphic to the other, because the logged events are different at each of the sub-datasets. As a result, the graphics reveal an evolution in the similarity values over time. An example of this technique was created with four time slots. Such a value was arbitrarily selected because it provides sub-datasets with a duration of about one month, and the result is easily understandable. It is shown in Figure 3. This visualization was developed with the goal of understanding how time affects the visualization, and was not used in the validation with users.

2.3 Use of the Visualization Technique to Make Score Estimations

The graphics shown in the previous section are based on a complete knowledge of the whole of the assessment of a student during the course. In other words, these are the graphics obtained at the end of the semester. However, the initial intent is to inform students about their trajectory *during* the semester without waiting until the final exam has been passed. Regardless of this limitation, the technique to create these graphics has been refined to provide feedback during the course.

The idea is to compare the behavioural similarities of the students on the current course with those from previous course editions. That is, if two editions of a course follow the same learning flow, the reference student is selected from those in the on-going course, while the rest of the dataset belongs to students from the previous course editions. All the points on the graphic belong to the previous edition's students, so their scores (their positions) are known and they can be drawn. The green intensity of the drawn points is given by their similarity to the reference student. Thus, the student will be able to know how similar students performed in previous courses, and such information will increase their awareness and promote self-reflection.

In order to validate the proposal, we needed the data of students of previous courses, and also data of students in a currently running course. In order to simulate such a situation with a single dataset, the selected reference student was treated as a "student of the current course" and the rest of the students were treated as "students of the previous courses". This method applies for the validation with users explained in Section 5.

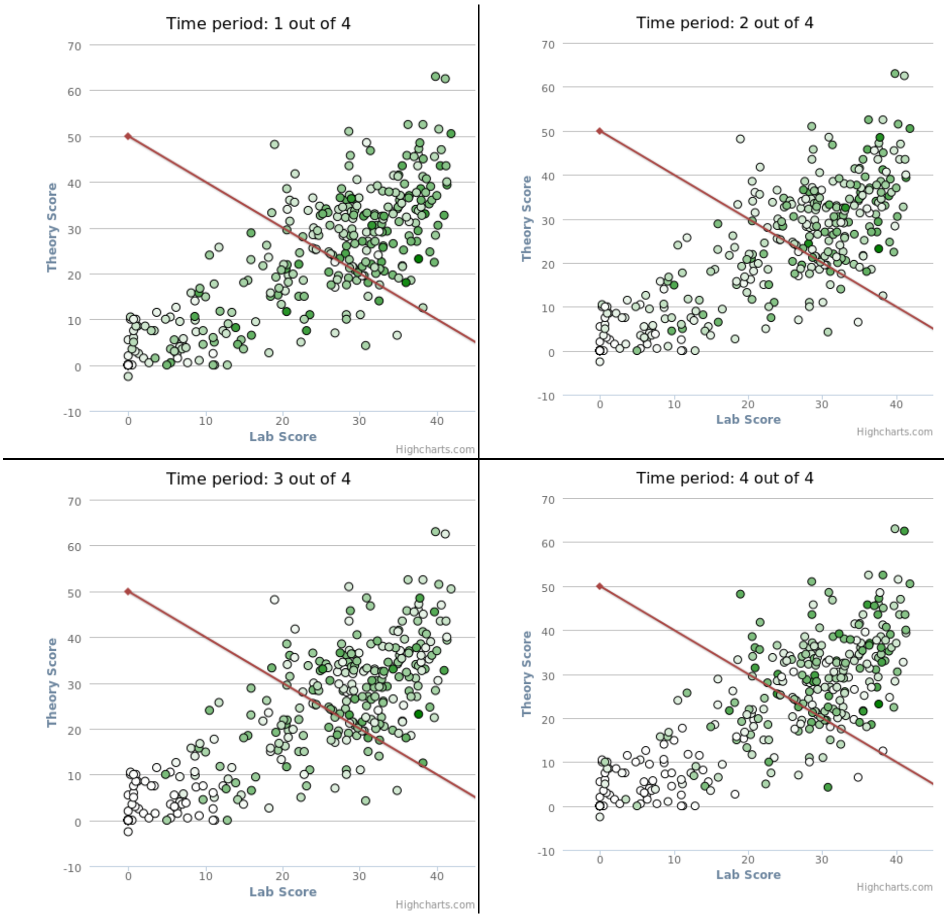


Figure 3: Visualizations for different time periods

3 Similarity Measurement

In the proposed graphics, each point represents one student and the intensity of its colour represents her similarity to the reference student. This section details how such similarity is calculated.

3.1 Translation of Event Logs Into Similarity Measures

In a typical case, the dataset created by tracking the students' activities is derived from a set of log files where each line represents an event produced by one person at a particular moment. In order to perform a similarity measurement, the initial data needs to be translated into vectors that are suitable for being processed mathematically. This subsection explains such pre-processing steps, and demonstrates the process with a small log file for the sake of clarity. The example log is shown in Listing 1. Each row in the logs is a succinct representation of events such as "*student Alice views the forum*" or "*student Charlie opens the compiler*".

```
Timestamp, type, studentId
2011-11-22, lms_view_forums, alice
2011-11-23, lms_view_forums, bob
2011-11-23, lms_view_forums, charlie
2011-11-23, bashcmd, alice
2011-11-24, bashcmd, charlie
2011-12-01, bashcmd, bob
2011-12-02, lms_view_forums, alice
2011-12-02, gcc, charlie
2011-12-02, bashcmd, alice
2011-12-03, gcc, charlie
2011-12-03, bashcmd, bob
2011-12-03, bashcmd, alice
```

Listing 1: Example of event logs

In order to calculate the similarities, the first step is to translate the event log into vectors that represent the students' activities. This is achieved by counting the number of events of a certain type produced by a user. The process is repeated for each pair event/student, resulting in a matrix like that presented in Table 1, constructed with the data from Listing 1. In the matrix, each row represents a different student, and the columns represent different event-types. As such, each cell contains the number of times that a given student has produced a certain event-type.

	lms_view_forums	*bashcmd*	*gcc*
alice	2	3	0
bob	1	2	0
charlie	1	1	2

Table 1: Number of events per user, extracted from the example

With the students' data tabulated and with each row representing a student as a vector, the next step is to calculate the similarity between two vectors. We surveyed several mathematical methods for the similarity measurement. For the purpose of this work, Bray-Curtis similarity [Bray & Curtis, 1957] has been applied, as it yielded the best classification results (see Section 3.2 for details). This measurement method was initially proposed in ecological studies to quantify the difference between two samples of ecological abundance taken at different locations.

Formally, for the similarity between two students (i and i'), with the events number of student i and type j denoted by n_{ij} , and the totals (the sum of the row of values) as n_{i+}, the Bray-Curtis similarity is given by Equation 1:

$$b_{ii'} = 1 - \frac{\sum_{j=1}^{N} |n_{ij} - n_{i'j}|}{n_{i+} + n_{i'+}}$$

Equation 1: Bray-Curtis similarity for students i and i'

One advantage of this method is that the output values are within the range of 0 to 1. A value of 0 means 'completely disjoint samples' while 1 means ‘identical samples’. The result is a symmetrical matrix with the cell i, i' containing the similarity between students i and i'. An example is given in Table 2:

	Alice	***Bob***	***Charlie***
Alice	0	0.75	0.44
Bob	0.75	0	0.57
Charlie	0.44	0.57	0

Table 2: Bray-Curtis similarities for the sample data

The inclusion of temporal information in this context has been proved to be an effective approach in similar areas, such as recommendations [Campos, Díez & Cantador, 2013], but the presented similarity measure discards such information (i.e., the event time-stamp is not considered in the determination of similarity values). To address this issue, the graphics can be constructed for an arbitrary time period by simply restricting the analysis to those logs within this time period.

3.2 Selecting the Best Similarity Measure

The goal of this research is to serve as a proof of concept of the visualization method. That is, to determine if the presented visualization method can result in a graphic where student similarities are used to foresee student scores. Therefore, it is out of the scope of this research to find the best performing metric, and this is part will be left for future work.

We surveyed the *Euclidean distance, Euclidean-squared distance, Pearson Correlation Coefficient, Spearman’s Rank Correlation Coefficient and Bray-Curtis Similarity* methods. In fact, some of these methods are distances and not similarities. In order to always work with normalized similarity (i.e., metrics where 0 means 'completely disjoint samples' while 1 means ‘identical samples’), we have normalized the results so the output of each metric is always a number between 0 and 1 that complies with the above condition.

This survey is a first approach in determining which of these methods (if any) provides a graphic in which similar students are located close to one another (i.e., whether students who behave in a similar manner are graded similarly). Two approaches were considered in order to select the best measure: visual inspection and root mean squared error (RMSE) calculation.

The *visual inspection* simply consisted of a manual revision of several graphics in order to subjectively determine whether a person would be able to recognize a cluster in the graphic. In a joint discussion, the researchers agreed on designating 'Bray-Curtis' as the best measurement method for this purpose.

For the *RMSE* calculation approach, the students' scores were estimated as a weighted average of their peers' scores, with weights given by the normalized similarity (i.e., from 0 to 1). Next, the RMSE was calculated for each of the estimators, resulting in the values presented in Table 3. The values show that all the estimators performed better during the final part of the course (time slot four), and that the best results were offered by Bray-Curtis similarity.

The authors acknowledge that the validity of the selection method is restricted to this particular scenario and dataset. Future work will include the definition of a more systematic methodology to determine, among a wider set of metrics, which is the most suitable metric for a certain scenario.

Estimator		**RMSE**			
	Time slot:	**1/4**	**2/4**	**3/4**	**4/4**
Euclidean		26,38	25,97	25,25	24,56
Euclidean-squared		26,12	25,58	25,19	24,39
Pearson		27,05	26,59	25,78	25,25
Spearman		26,99	26,60	25,59	25,21
Bray-Curtis		26,91	25,57	22,95	21,99

Table 3: RMSE calculated for each of the estimators in the different time slots

Both the visual review and the RMSE calculation suggested that the Bray-Curtis similarity offers the best results, and therefore this was the method used in the empirical validation.

4 Research Methodology

The presented work followed a two-step research methodology. In the first step, a use case deployed in a real educational setting offered a complete dataset that logged the students' activities and scores over weeks and, after the finalization of the course, the proposed visualization technique was built and tested using the collected dataset. In the second step, an experiment was deployed in order to empirically validate the research hypothesis. This section presents these two methodological steps as separate subsections.

4.1 The Use Case and Monitoring Mechanism

The use case was deployed in a second year engineering course at a higher education institution during the fall semester of 2010. The course followed an active learning strategy based on the *flipped classroom approach*: for each session, a set of previous activities was defined and the students were supposed to work in these activities before attending the session in which this material was reviewed and expanded. Among the learning outcomes expected from the course, one of them mentioned explicitly the proficient use of certain tools. As a consequence, completing the course implied taking on a significant load of practical work that the students carried out during their personal study time. This subsection depicts the pedagogical and technical characteristics of the use case, while its complete description falls outside the scope of this work. Further details can be found in other published studies [Pardo, Kloos, 2011].

The data was obtained by trying to maximize the observation capability and extending it beyond the limits of the LMS. In highly practical learning experiences, most of the relevant interactions occur outside the LMS. The approach adopted to capture the dataset was based on the use of virtual appliances [refs-removed-for-blind-review]. The students were given a fully configured virtual appliance suitable for installation on their personal computers. This appliance contained all the applications and environmental configurations required to participate in the course activities. Among these applications was a software version control program used to exchange documents with a central server. This application, combined with the requirement to work on a collaborative project in teams of four members, translated into the frequent exchange of documents. The space shared by the teams was also available to the teaching staff.

The applications installed in the virtual appliance were instrumentalized so that every time they were invoked, they recorded the date/time of the event as well as some additional execution parameters. This instrumentalization was done on a per-tool basis. In general, the process amounted to writing a *wrapper* application that recorded the information before invoking the genuine program.

The information collected while the students used these applications needed to be relayed back to a central server so that it could be further processed. This transmission was performed using the data exchange with the software version control application. Every time the user submitted a new change to that repository, an instrumentalized version of the application added the recorded events as part of the payload of the transaction. With this technique, several aspects of the capturing mechanisms were achieved. The first one was that events were communicated through a secure link following an authentication step (i.e., imposed by the software version control application). The second was that the recorded information became part of the shared space of the user and, therefore, was readily available to them to manage. The third one was that the software control version application offered the historical storage of changes in these files, thus allowing the recovery from potentially catastrophic failures in either the student environment or the central server.

Thus, while students worked on their assignments and exchanged documents with their peers, the system collected a set of logs that were unequivocally attached to a user and which contained in each line the name of the event that had occurred (i.e., the platform that was invoked), a date/time stamp of when the event occurred and, in some cases, additional parameters of the event, such as the file name that was edited.

Privacy issues were addressed by informing the students about the collection mechanism, obtaining their consent and, more importantly, by providing full access to the data and the possibility of deleting the collected information at any point in time. The virtual appliance was also equipped with a simple mechanism to disable the instrumentalization at any point in time.

The monitoring mechanism was deployed during a 15-week period from September to December 2010. From a total of 248 students enrolled on the course, events were recorded for 172 of them (a total of 446,418 events). However, of the 76 students for which no event was recorded, 30 participated in the course activities while the rest dropped from the course.

4.2 Validation with Users

The proposed graphics were created with the data collected at the above explained use case. Afterwards, a validation with practitioners (teachers) was set up in order to determine the validity of the visualization technique. The remainder of this section is devoted to explaining the empirical validation, including the addressed research questions and the research methodology.

4.2.1 Research Questions

The proposed visualization technique does not provide an estimation of the score, in the sense that it does not provide a number. Instead, the idea is to show structured information to the student so that she can guess the score. An experiment was conducted in order to determine whether or not the graphics actually enabled human estimations as well as the accuracy of such estimations. In short, the experiment was guided by the following research questions:

[RQ1] How accurate is the method when used to estimate failure/success?
[RQ2] How accurate is the method when used to estimate the numerical score?

A secondary goal of the experiment was to detect any possible usability issues of the proposed visualization technique. That is, the volunteers were also surveyed regarding the following research questions:

[RQ3] How easy is it to interpret the graphics?
[RQ4] What is the perceived usefulness of the graphics?

4.2.2 Experimental Setting

A total of 11 volunteers participated in the experiment at different moments. They were practitioners (teachers) who were unfamiliar with the presented research, and who did not take part in the course that generated the dataset. Therefore, they had to first attend a training session in order to understand the visual representation and its goal. Next, they received some examples of the graphics (each of them with a randomly selected reference user) and, for each of the examples, they were asked to make an estimation of the reference users' total score. The whole session lasted about 20 minutes per volunteer. Despite that the visualization technique was targeted to students, it had to be first validated as a proof of concept, and this needed to be done with volunteers. This fact may have biased the results, as discussed in Section 5.4.

The training session began with a verbal description of the visualization technique, which contained some graphical examples. At this step, a researcher was present and answered the volunteers' questions. Afterwards, and while still in the training session, the volunteers solved an exercise in which they reviewed some graphics with the reference student - randomly selected - marked as a red diamond (as with the example in Figure 1), and they answered whether they considered the red diamond to be, taking into account the other students' scores, predictive of their own performance.

After the training session, the actual experimental setting began and the volunteers reviewed a number of different graphics in which the reference student was not drawn (as with the example in Figure 2). They were asked to make an estimation of the hidden student's total score in different ways: by guessing whether or not the student had passed the course, guessing the numerical score and, finally, selecting the area in the graphic where, in their opinion, the hidden student was actually located. In these questions, NA (not available) was allowed in case the participants did not have enough information to guess the total score. Finally, and after several repetitions of this step, the volunteers answered some general questions regarding the usability of the graphics. A printable version of the questions posed to the volunteers is provided in Appendix A.

4.2.3 Data Analysis Method

Each of the four research questions stated in Subsection 4.2.1 has a different nature, and therefore has been analysed from a different perspective. The first two questions regard the accuracy of the estimations, and so their analysis followed a quantitative approach.

More precisely, the analysis of RQ1 is based on the percentages of successful estimations under different circumstances. The answer to RQ2 requires a more complex analysis, with some different methods that complement each other. This includes the following:

- Analysis of the error values, including average error, standard deviation, significance tests and a discussion of the histogram of error values.
- Balance of overestimations and underestimations, with the aim of determining whether or not the graphics force the user to make estimations that are greater or lower than the actual value.
- Success rate in the attempts to locate the reference student in the graphic.

On the other hand, the answers to RQ3 and RQ4 are a secondary goal of the experiment, and the collected data does not offer a sufficiently large sample for the statistical analysis. Therefore, their analysis was confronted from an exploratory point of view, and the conclusions will therefore serve as a guide for further research. In particular, their analysis includes a discussion of the answers to the final questionnaire.

5 Results and Discussion

This section presents and discusses the results of the empirical evaluation. The discussion is guided by the research questions.

5.1 Failure/Success Estimation

The accuracy while estimating failure/success (RQ1) can be viewed from several points of view, as presented in Table 4.

The first approach (the first row in the table) is to measure the percentage of successful estimations over all the obtained answers. The result (67.27%) is positive, but far from our initial expectations. The measure includes all the answers, including ‘NA’; however, ‘NA’ answers represent those cases where the volunteer did not perform an estimation. The second row in the table presents the success percentage when ‘NA’ answers are not considered. Taking into account that this research is a proof of concept and that the parameters for the similarity measure allow for a finer tuning, we think that 74% for successful estimations is a strong result.

	Measure	**Percentage**
1	Successful estimations over all the answers (including NA)	67.27%
2	Successful estimations over non-NA answers	74.00%
3	Successful estimations over those who passed (excluding NA)	90.00%
4	Successful estimations over those who failed (excluding NA)	50.00%
5	“Say-pass” successful estimations	72.97%
6	“Say-fails” successful estimations	76.92%

Table 4: Percentage of success in pass/failure estimations

Table 4 also presents the percentage analysis from two different perspectives:

- Estimations classified by the actual results of the students (rows 3 and 4). That is, we separated the students into two groups (those who passed and those who failed) and analysed the percentages of successful estimations in these groups. The accuracy is clearly better among those students who passed while it is below expectations when they actually failed.
- Estimations classified by the given prediction (rows 5 and 6). That is, we separated the estimations into two groups: those that predicted success and those that predicted failure. In this case, the chances of success in the estimation are a little bit higher when predicting failure, but the percentages are pretty similar.

In summary, the chances of success in pass/failure estimations validate the proposed visualization technique as a proof of concept, but further research and the adjustment of the similarity measurement and borderline cases are required.

5.2 Numeric Score Estimation

The accuracy of the score estimations was analysed from two points of view. First, the volunteers were asked to estimate the numerical score of the hidden student. Second, they were asked to select with a mouse the rectangle in which they would locate the student. This section presents the results of these two approaches.

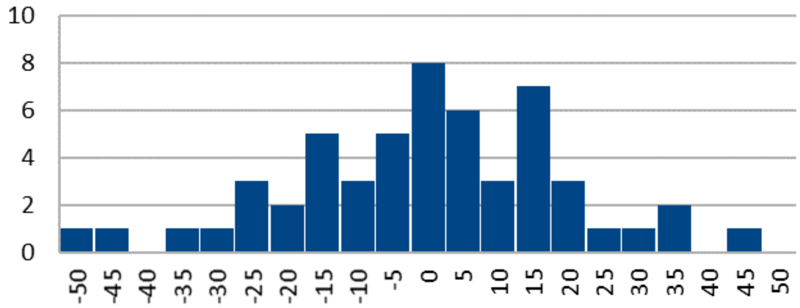


Figure 4: Histogram of error in estimations

Mean:	-1.92
Standard deviation	19.36
95% confidence interval for the mean	[-7.20, 3.36]
Skewness coefficient	-0,26
Shapiro-Wilk test p-value	0.9211

Table 5: Descriptive statistics for the error in the estimations

The average error value is near zero, clearly showing that the graphics allowed for the estimation of the students' scores. However, the dispersion of the samples remains too high. For example, 60% of the estimations had an error lower than 15, 78% lower than 20 and 86% lower than 25. In other words, the estimations given by the volunteers did not ensure any specific result, but were sufficiently valid to have a rough idea of the expected student's performance.

Another interesting parameter for analysis is the *skewness coefficient*. This value shows whether the distribution "leans" towards positive or negative values. The obtained value, near zero, means that overestimations are as likely as underestimations.

Table 6 summarizes the analysis of the second method to measure the estimations accuracy, which is the error in area selections. Here, a volunteer succeeded in her estimation if the selected area actually contained the hidden students, which happened in 26 cases (47.27%). However, some of the selected areas were too large; that is, some of the volunteers selected an area that included almost all of the points in the graphic, and it obviously included the hidden student. If we exclude from this analysis those area selections with a range (max. score – min. score) greater than 30, the success rate was a discrete 30%. This result indicates a low accuracy, which contrasts with the analysis of the numeric score estimations. This suggests that it is easier to have an intuitive idea of the hidden student's score than it is to actually locate her in the graphic.

Successful area selections	47,27%
Area selections with range < 30	72,73%
Successful area selections with range < 30	30%

Table 6: Success percentages in area selection

5.3 Ease of Use and Perceived Usefulness

The proposed visualization technique aims at raising student awareness, and for this purpose they should be able to interpret the graphics with little effort. Thus, the final questions posed to the volunteers attempted to explore the perceived ease of use and usefulness of the graphics. A Likert scale was used, where 1 meant completely disagree and 5 meant completely agree. Table 7 summarizes the results of those questions, answered. Due to the limited number of available answers, it is not possible to claim statistical significance of these conclusions. The analysis of ease of use and perceived usefulness are studied here from an exploratory point of view, and further research will be required to confirm the results.

The first two questions are related to the ease of use of the graphics. The collected answers suggest that the volunteers were able to interpret the graphics, but that their interpretations were not straightforward. They thought that the students would be able to understand the graphics, but not at first glance. This idea aligns with the confidence expressed while doing the estimations (see Table 8). Future work should be devoted to improving the usability and understandability of the proposal.

The perceived usefulness addressed in questions three, four and five had a positive evaluation, and most of the volunteers thought that the graphics would actually provide useful self-awareness for the students. They also considered (although with less consensus) that such self-awareness would promote self-reflection and, therefore, have an impact on students' behaviour. Moreover, the volunteers considered this to be a useful tool that would be used by the students.

In summary, the collected answers suggest a good result for the perceived usefulness of the graphics, but their ease of use did not match the researchers' expectations.

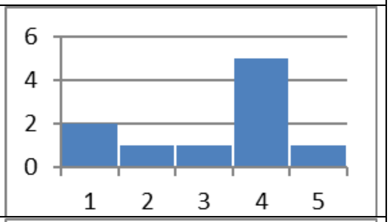
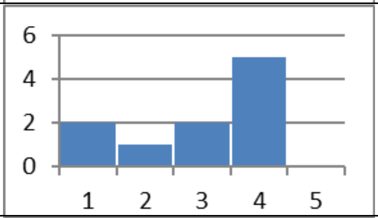
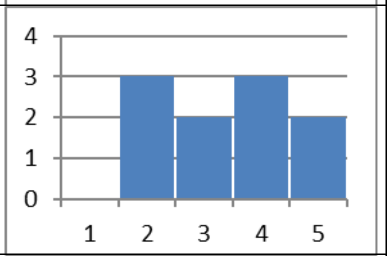
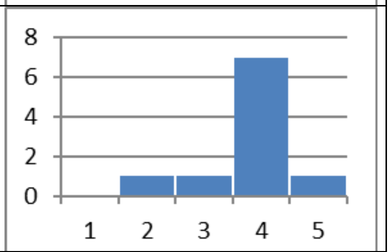
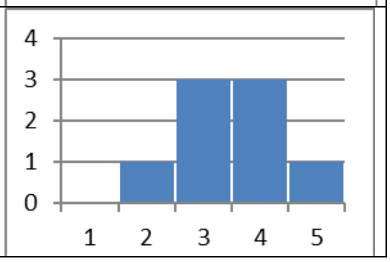
	Question	Mean	Std	Histogram
1	I found the graphical representation easy to interpret	3.2	1.40	
2	I think that the students will learn to interpret the graphics very quickly	3	1.25	
3	I think that the graphics will have an impact on students' behaviour	3.4	1.17	
4	I think that the students will find the graphics useful in gaining self-awareness	3.8	0.79	
5	I think that the students would like to use the system frequently	3.5	0.93	

Table 7: Analysis of ease of use and perceived usefulness

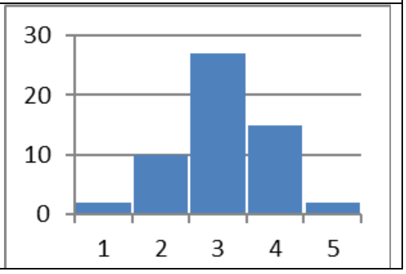
Question	Mean	Std	Histogram
How confident with your estimation are you?	3.09	0.86	

Table 8: Confidence level in the volunteers' estimations

5.4 Limitations of the Study

The researchers recognize certain limitations in the methodology that could have biased the study. However, in the researchers' opinion, these limitations do not invalidate the analysis. Below is a description of these possible issues:

- Students are the target users of the presented visualization technique. However, due to organizational reasons, the volunteers that participated in the experiment were practitioners and not students. In this study, the researchers' intention was to determine whether a person is able to make score estimations, and they concluded that a survey with practitioners can reach this goal. Furthermore, they considered it appropriate to first refine metrics and explanations before presenting the visualization to the students. The current study is a proof of concept, and future research will aim at experimentation in authentic learning scenarios.

- The green and red colours used in the graphics do not take into account the limitations of colour-blind people. One of the volunteers was recognized as suffering such a deficiency, but the analysis presented does not take into account this fact. Further conversations with this volunteer led to the conclusion that this factor did not affect the results, but future studies might be devoted to confirm or reject such an assertion.

- Some of the volunteers complained about the area selection method. In particular, they said that the rectangular shape of the selection did not allow them to include within the selection all the areas that they wanted. This fact might have affected the results discussed in Section 5.2.

6 Conclusions and Future Work

This article presents two outcomes in combination: a visualization technique based on similarity measures, and a technique to estimate the final score of students in a course. The visualization technique is produced as a method for providing self-awareness and for fostering self-reflection. To this extent, the technique makes it easy for students to estimate their score for a course. The visualization makes use of the concept of 'similarity', and this method is based on the idea that the students' final grades seemed to be similar if their behaviour looked alike. The empirical evaluation presented in this article provides a proof of concept for this idea, subject to ongoing research and awaiting deployment in practical scenarios with real students.

This visualization concentrates on the similarity between the students' behaviour, that is, a single figure that summarizes a number of the activities and assessments carried out by the student so far. In the context of a student group, the combination of single values by every student produces a graphical representation in the form of a trajectory. This trajectory is easily understandable by users, and becomes a mean for predicting their own progress and final grade.

The suggested visualization was evaluated with 11 users in order to determine whether they were able to estimate the students' score by looking at the graphic alone. The analysis of the results showed that the visualization is valid in providing an idea of the students' performances. An exploratory analysis of the perceived usefulness suggests that the visualization will increase the students' awareness, and therefore will have a clear positive impact on their behaviour. In other words, the visualization of the projection of one's own score, in comparison with those of similar users, suggests an increase in self-awareness, and this might lead to better performance.

The complete understanding of the proposed visualization required a training session with the users so that they could provide meaningful input. In this experimental setting, we have designed a brief web-based tutorial and the results suggest that the low amount of provided user experience is a weakness of the proposal. Certainly, one of the guidelines for future work would be the creation of a multimedia tutorial that helps in understanding the interface and the basic use of the tool, so that it can have clear and positive implications on students' awareness and progress. An additional main guideline for future research would concentrate on the accuracy of the estimations. This task involves further research and fine-tuning of the similarity measurement methods, which would allow for more accurate score estimations in addition to a better understanding of the visualization technique itself.

Acknowledgements

This work has been partially funded by the INTUITEL project (FP7/2007-2013, grant agreement: 318496) and EEE: Orquestando Espacios Educativos Especulares, ``Plan Nacional de I+D+I'' (TIN2011-28308-C03-03).

References

[Baker & Yacef, 2009] Baker, R. S. J. D., & Yacef, K. The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, *1*(1), 2009, 3–17

[Binkley et al., 2010] Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., & Rumble, M. Draft White Paper 1: Defining 21st century skills. *Melbourne: ACTS*. 2010.

[Bray & Curtis, 1957] Bray, J. R., & Curtis, J. T. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, *27*(4), 1957, 325–349.

[Campbell et al., 2007] Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. Academic Analytics. *Educause Review*, *42*(October, 2007), 40–57.

[Campos, Díez & Cantador, 2013] Campos, P. G., Díez, F., & Cantador, I. Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols. *User Modeling and User-Adapted Interaction*.2013

[Carmean & Mizzi, 2010] Carmean, C., & Mizzi, P. J. The Case for Nudge Analytics. *EDUCAUSE Quarterly*. 2010

[Cocea & Weibelzahl, 2009] Cocea, M., & Weibelzahl, S. Log file analysis for disengagement detection in e-Learning environments. *User Modeling and User-Adapted Interaction*, *19*(4),2009, 341–385.

[Davis, 1989] Davis, F. D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 1989, 319.

[Donovan et al., 2008] Donovan, J. O., Smyth, B., Gretarsson, B., Bostandjiev, S., & H, T. PeerChooser : Visual Interactive Recommendation. *Computer Science and Informatics*, 2008, 1085–1088.

[El-Bishouty et al., 2007] El-Bishouty, M. M., Ogata, H., & Yano, Y. Personalized Knowledge Awareness Map in Computer Supported Ubiquitous Learning. *Sixth IEEE International Conference on Advanced Learning Technologies ICALT06*, *10*(4), 2007, 817–821.

[Ertmer & Newby, 1996] Ertmer, P. A., & Newby, T. J.. The expert learner: Strategic, self-regulated, and reflective. *Instructional Science*, *24*(1), 1996, 1–24.

[Ferguson, 2012] Ferguson, R. *The State of Learning Analytics in 2012 : A Review and Future Challenges a review and future challenges*. 2012

[Glahn et al., 2009] Glahn, C., Specht, M., & Koper, R. Visualisation of interaction footprints for engagement in online communities. *Educational Technology & Society*, *12*(3), 2009, 44–57.

[Govaerts et al., 2012] Govaerts, S., Verbert, K., Duval, E., & Pardo, A. The student activity meter for awareness and self-reflection. *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA 2012* (p. 869). New York, New York, USA: ACM Press.

[Keim et al., 2008] Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., & Melançon, G. Visual Analytics: Definition, Process, and Challenges. In A. Kerren, J. T. Stasko, J.-D. Fekete, & C. North (Eds.), *Information Visualization* 2008, (pp. 154–175). Berlin, Heidelberg: Springer Berlin Heidelberg.

[Larusson & Alterman, 2009] Larusson, J. A., & Alterman, R. Visualizing student activity in a wiki-mediated co-blogging exercise. *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems CHI EA 2009*, 4093.

[Lee et al., 2003] Lee, M. D., Butavicius, M. A., & Reilly, R. E. Visualizations of binary data: A comparative evaluation. *International Journal of Human-Computer Studies*, *59*(5), 2003, 569–602.

[Leony et al., 2012] Leony, D., Pardo, A., De la Fuente Valentín, L., De Castro, D. S., & Kloos, C. D. GLASS. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK 2012* (p. 162). New York, New York, USA: ACM Press.

[Lew & Schmidt, 2011] Lew, M. D. N., & Schmidt, H. G. Self-reflection and academic performance: is there a relationship? *Advances in health sciences education : theory and practice*, *16*(4), 2011, 529–45.

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

[Mazza & Dimitrova, 2004] Mazza, R., & Dimitrova, V. Visualising student tracking data to support instructors in web-based distance education. *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters* (pp. 154–161). 2004, ACM.

[Mazza & Dimitrova, 2007] Mazza, R., & Dimitrova, V. (2007). CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses. International Journal of Human-Computer Studies, 65(2), 125–139.

[McCrindle & Christensen, 1995] McCrindle, A. R., & Christensen, C. A. The impact of learning journals on metacognitive and cognitive processes and learning performance. *Learning and Instruction*, *5*(2), 1995, 167–185.

[McGrath, 2010] McGrath, O. Data Mining User Activity in Free and Open Source Software (FOSS)/ Open Learning Management Systems. *International Journal of Open Source Software and Processes*, *2*(1), 2010. 65–75.

[Munoz-Organero et al., 2010] Munoz-Organero, M., Munoz-Merino, P. J., & Kloos, C. D. Student Behavior and Interaction Patterns With an LMS as Motivation Predictors in E-Learning Settings. *IEEE Transactions on Education*, *53*(3), 2010, 463–470.

[Pardo, Kloos, 2011] Pardo, A., & Kloos, C. D. Stepping out of the box. Towards analytics outside the Learning Management System. *International Conference on Learning Analytics, 2011*.

[Potgieter et al., 2010] Potgieter, M., Ackermann, M., & Fletcher, L. Inaccuracy of self-evaluation as additional variable for prediction of students at risk of failing first-year chemistry. *Chemistry Education Research and Practice*, *11*(1), 2010, 17.

[C. Romero & Ventura, 2007] Romero, C., & Ventura, S. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, *33*(1), 2007, 135–146.

[Romero & Ventura, 2011] Romero, Cristóbal, & Ventura, S. Preface to the special issue on data mining for personalised educational systems. *User Modeling and User-Adapted Interaction*, *21*(1-2), 2011, 1–3.

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

[Romero-Zaldivar et al., 2012] Romero Zaldívar, V. A., Pardo, A., Burgos, D., & Delgado Kloos, C. Monitoring Student Progress Using Virtual Appliances : A Case Study. *Computers & Education*, *58*(4), 2012, 1058–1067.

[Santos et al., 2013] Santos, J. L., Verbert, K., Govaerts, S., & Duval, E. Addressing learner issues with StepUp! *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK 2013* (p. 14). New York, New York, USA: ACM Press.

[Siemens, 2012] Siemens, G. Learning analytics: envisioning a research discipline and a domain of practice. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK 2012* (p. 4). New York, New York, USA: ACM Press.

[Verbert et al. 2013] Verbert, Katrien, et al. "Learning analytics dashboard applications." American Behavioral Scientist (2013): 0002764213479363.

[Verpoorten et al., 2011] Verpoorten, D., Westera, W., & Specht, M. A first approach to "Learning Dashboards" in formal learning contexts. 1st International Workshop on Enhancing Learning with Ambient Displays and Visualization Techniques at EC-TEL 2011, Palermo, Italy.

[Wang et al., 2011] Wang, M., Peng, J., Cheng, B., Zhou, H., & Liu, J. Knowledge Visualization for Self-Regulated Learning. *Educational Technology and Society*, *14*, 2011, 28–42.

Appendix A

This appendix shows a printable version of the questions posed to the volunteers.

Please answer the following questions:

Does the reference student pass the course? -

In which area of the graphic is the refence user positioned? Select the area with the mouse. Your selected area:

Select your guess for the final grade of the reference student. -

How confident with your estimation are you? NA 1 (Completely unconfident) 2 3 4 5 (Completely confident)

Please indicate to what extent do you agree or disagree with the following statements:

	NA	1 (Strongly Disagree)	2	3	4	5 (Strongly Agree)
I think that the students would like to use the system frequently						
I found the graphical representation easy to interpret						
I think that the students will learn to interpret the graphics very quickly						
I think that the graphics will have an impact on students' behaviour						
I think that the students will find the graphics useful to get self-awareness						

Figure 5: printable version of the questions posed to volunteers