D3.1 Framework of Quality Indicators

Learning Analytics Community Exchange
www.laceproject.eu

Framework of Quality Indicators

Public Deliverable – D3.1

Deliverable Coordinator: Maren Scheffel, Hendrik Drachsler
Coordinating Institution: Open University of the Netherlands (OUNL)

Data due: 31.10.2014 - **Submission:** 31.10.2014

Keywords: group concept mapping, quality indicators, framework, tool, evaluation study

Abstract:
This document presents results from a group concept mapping study that was conducted with learning analytics experts to construct a framework of quality indicators for learning analytics. This framework was then turned into an applicable tool to measure those quality indicators and evaluated by project members.

Contents

Introduction

Making use of learning analytics (LA) can give added value to both learners and educators. LA can help learners to plan and reflect these activities better by becoming aware of their actions and learning processes (see Endsley, 1995, 2000; Schön, 1983). Reflection can promote insight about something that previously went unnoticed (Bolton, 2010) and lead to a change in learning behaviour. Results of LA can be used to foster awareness and thus reflection or to give recommendations for further steps in a current learning scenario. As Ferguson (2014) explains, LA offers "ways for learners to improve and develop while a course is in progress. These analytics do not focus on things that are easy to measure. Instead, they support the development of crucial skills: reflection, collaboration, linking ideas and writing clearly". Awareness and reflection support for students are consequently highly important aims of learning analytics. Progress toward these objectives and their impact, however, are hard to measure due to the lack of standards that the student support of LA tools can be measured against.

The same applies in the case of educators. In order to support students on a course, teachers should be aware of what the students are doing, how they are interacting with the course material, and where comprehension problems arise (cf. Scheffel et al., 2011; Scheffel et al., 2012). Zinn & Scheuer (2006) conducted a survey among teachers, trying to identify requirements for student tracking tools. Among the information deemed most important were: the students' overall success rate, their mastery level of concepts, skills, methods and competencies, and the most frequently diagnosed mistakes. Such information is also needed for the evaluation of elements of a course, including pedagogy, materials, contents, tools, and tests. Awareness and reflection support for educators are thus also highly important aims of LA. But as in the case of learner support, standards that define quality indicators for learning analytics tools are missing.

While the potential added value of LA for learners and educators is clearly recognised (Long & Siemens, 2011), little research has been done so far to compare the findings of empirical LA studies in order to investigate which analytics and analytic tools have a desirable effect on learning. We have therefore worked towards the identification of quality indicators for learning analytics that will help standardise the evaluation of LA tools. Drawing on the experience of Drachsler et al. (2014) who have already successfully created an evaluation framework specifically for data competitions in TEL, we have conducted two studies for the development of an evaluation framework of quality indicators for learning analytics.

The quality indicator framework has been developed with experts from the LA domain by using a Group Concept Mapping (GCM) approach (Scheffel et al., 2014). The GCM methodology as well as the study's results are presented in section 2, followed by a description of the first version of the framework in section 3. This framework is then turned into an applicable questionnaire and evaluated in section 4, followed by a conclusion in section 5.

Creating a Framework of Quality Indicators for Learning Analytics

Group Concept Mapping Methodology

Group Concept Mapping (GCM) is a highly structured approach that applies quantitative as well as qualitative measures in order to create a stakeholder-authored visual geography of ideas from a target group. These ideas are combined with specific analysis and data interpretation methods to produce maps that can be used to guide planning and evaluation efforts on the issues the group is considering (Kane & Trochim, 2007). In our study we used a GCM online tool (Concept Systems Global, 2014) with a three-step process: (1) generation of ideas, relating to quality indicators of learning analytics, (2) sorting of the collected ideas into clusters, and (3) rating of the ideas according to several values, including importance and feasibility. One important aspect of GCM is its bottom-up approach. Instead of presenting a given set of criteria to a group to sort and rate, the community itself generates the ideas that are to be clustered and rated by a group of experts. The individual input of the participants was aggregated by the tool to reveal shared patterns in the collected data by applying statistical techniques of multidimensional scaling and hierarchical clustering. Visualisations automatically produced within the tool then helped experts to grasp the emerging data structures and to interpret the data.

The involvement of participants in our GCM study was twofold. The first phase was conducted during the Learning Analytics and Knowledge Conference 2014 (LAK14). Calls for participation were circulated via several channels – including Twitter, project websites, personal contact and email – asking people involved and interested in LA to contribute their indicators for learning analytics to the brainstorming phase. Participation was accessible via a link that was open to all; there was no requirement to register with the GCM tool. In total, 74 people participated in the brainstorming phase.

For the second phase, sorting and rating of the collected quality indicators, we selected and made personal contact with 55 experts from the domain of LA. These were selected as experts on the grounds that they had been involved in the domain for several years, had published about learning analytics-related topics, were from the higher education sector and preferably had a doctorate. The demographics show that the average expert who engaged in the study was a researcher at a university with an advanced expertise in LA and with more than six or ten years of work experience.

Participants in all three activities were informed about the purpose, the procedure, and the time needed to complete the activities. Participants in the first phase were given a link to access the brainstorming section of the GCM tool and asked to generate ideas by completing the following statement: *"One specific quality indicator to evaluate the effects of learning analytics is …"*. Participants had ten days to contribute to the brainstorming. During this first phase, the 74 participants generated a total of 92 original ideas. Before using these statements in the second phase, identical statements were combined and overly vague ideas, such as 'Range of flexibility in moving from one point to another in a theoretical discussion', were removed. Statements that contained more than one idea were divided. For example, "students and teachers change their behaviour in some aspects" was split into one statement relating to teachers and another relating to students. After the data had been cleaned in this way, there were 103 statements (the full list is available at *http://bit.ly/103QILA*). These were randomised and moved into the sorting and rating phase.

During this phase, participants first sorted the statements according to their view of the statements' similarity in meaning or theme and began to name the resulting clusters. Dissimilar statements were not added to a "miscellaneous" cluster but rather into their own one-statement-cluster in order to ensure statement similarity within the clusters. Participants then rated all statements on a scale of 1 to 7 according to their *importance* and *feasibility*, with 1 being those with the lowest and 7 those with the highest importance/feasibility. The former refers to the priority or importance of a quality indicator in relation to the evaluation of effects of LA, and the latter indicates the perceived ease of applicability of a quality indicator. Participants had two weeks to complete the sorting and rating activities.

Results

From point map to cluster map

The GCM tool offers a number of automated analyses of the collected data: multidimensional scaling and hierarchical clustering for the sorting data and mean, standard deviation and correlation for the rating data.

Figure 1 shows a point map of the 103 statements, which is the outcome of the multidimensional scaling analysis. The multidimensional scaling analysis assigns a so-called bridging value between 0 and 1 to each statement according to the similarity to the statements around it. Statements with low bridging values have been grouped very close together with other statements around it, i.e. they are semantically closely related. For example, the group of statements on the lower right side of Figure 1 all deal with some form of student motivation and can be considered quite coherent. Statements with higher bridging values can also be grouped together but the surrounding statements are then further apart, e.g., statements 95, 23, 50, 61 – just below the centre of Figure 1 – all related to teacher motivation, engagement and feedback. Thus, statements that are close to one another in the map are also close to one another in meaning as they had been clustered together by the experts.

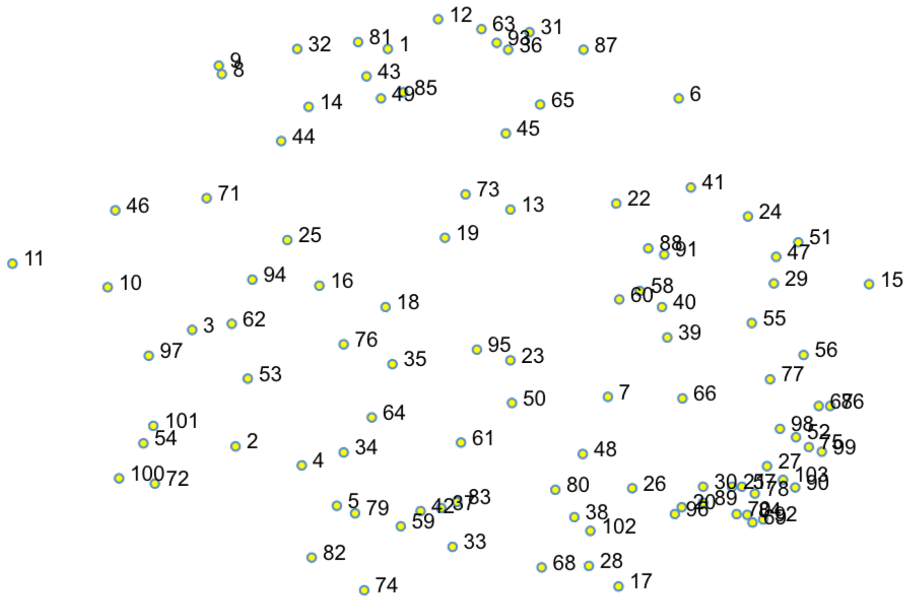

Figure 1: Point map of the 103 quality indicators

In some areas of the map it is quite easy to detect groups by simply looking at the point map. In other areas, however, it is more difficult to decide where group boundaries should be set. The hierarchical clustering analysis of the GCM tool offers several solutions to a given point map. We used a cluster replay map, starting with 15 clusters and working down to two (see Figure 2). For each cluster-merging step, we looked carefully at the statements within the clusters that were to be combined to check whether a merging made sense. The solution that the researchers judged best represented the collected data and the purpose of the study was the one with eight clusters.

After deciding on the number of clusters to work with, meaningful labels had to be assigned for the clusters. The system automatically suggests a list of labels per cluster. Another way of finding appropriate labels is to look at the bridging values of the statements within a cluster. The lower the bridging values are, the better those statements define the cluster. A third way to find meaningful cluster labels is to find the overarching theme of a cluster by looking at all statements within a cluster. We combined all three methods to define the labels of the eight-cluster solution (see Figure 3). This gave us the cluster labels *1. Data: open access*, *2. Data: privacy*, *3. Acceptance & uptake*, *4. Learning outcome*, *5. Teacher awareness*, *6. Learning performance*, *7. Learning support*, and *8. Student awareness*.

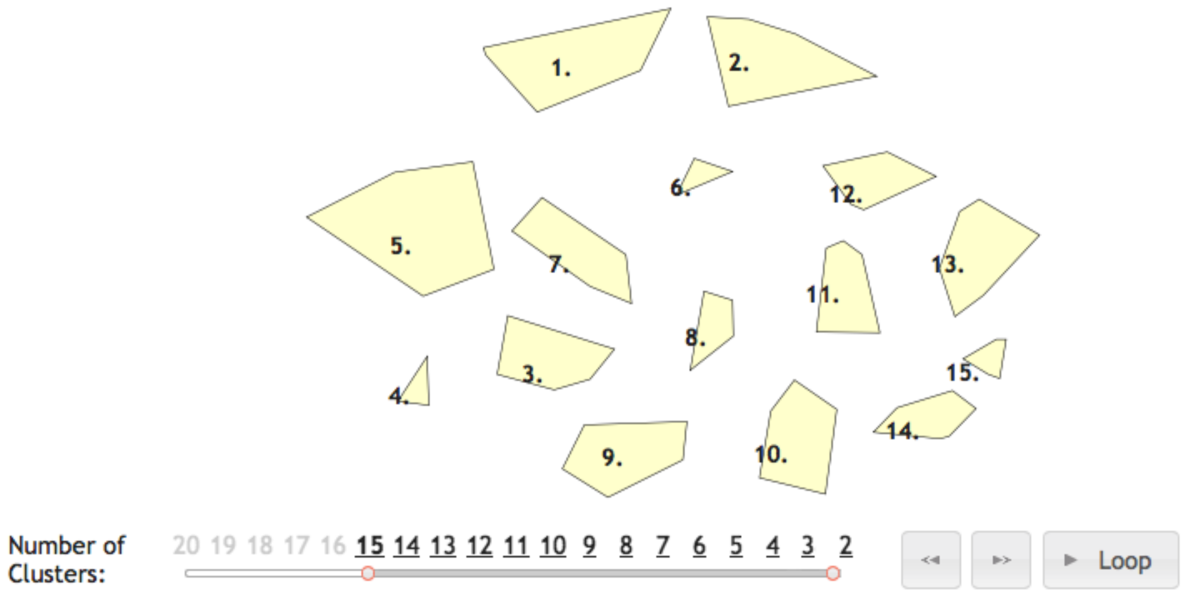

Figure 2: Replay map showing 15 clusters

The GCM tool assigns each cluster a bridging value. The more coherent a cluster is, the lower is its bridging value. A low bridging value can be attributed to a high agreement rate between the experts about statements within a cluster. The four most coherent clusters are *Student awareness* (0.11), *Data: open access* (0.25), *Data: privacy* and *Learning performance* (0.31 each). The clusters *Teacher awareness* (0.41), *Learning support* (0.45) and *Learning outcome* (0.46) are all similar in range. The cluster with the by far the highest bridging values and thus the least coherence is *Acceptance & uptake* (0.86). In order to increase understanding of the different clusters, a more detailed description of their characteristic statements is given below.

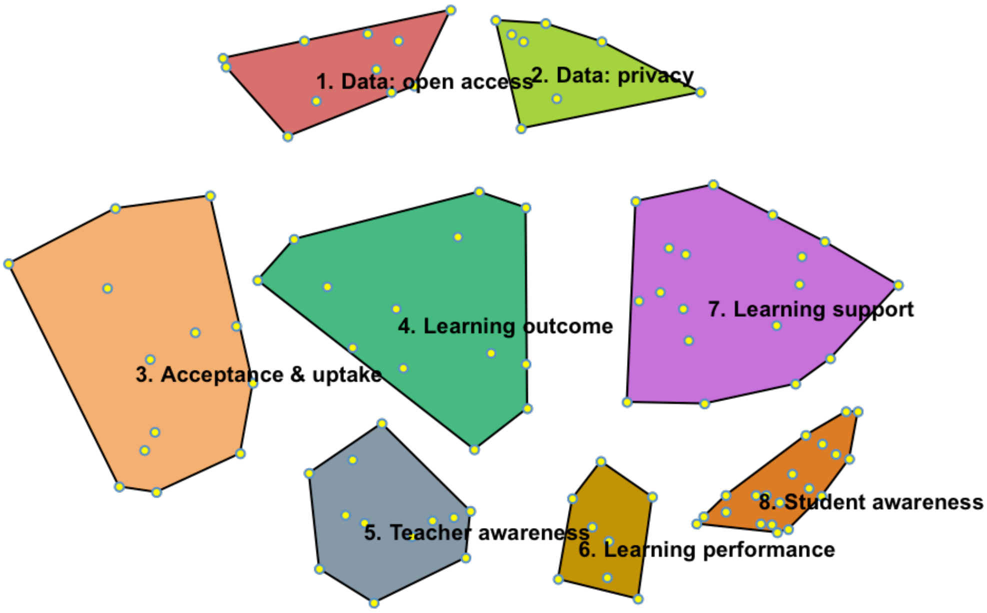

Figure 3: Cluster map with labels

Cluster 1 *Data: open access* (cluster bridging value (CBV) of 0.25) contains eleven statements with bridging values ranging from 0.06 to 0.60. Most of these statements deal with aspects of openness and transparency of data as well as the algorithms used, e.g., “that data are open access”, “portability of the collected data”.

Cluster 2 *Data: privacy* (CBV of 0.31) is about privacy, control of data, and transparency of data access. There are eight statements in the cluster with bridging values ranging from 0.10 to 0.72. Representative statements are “that privacy is ensured”, “if learners can influence which data are provided”.

Cluster 3 *Acceptance & uptake* (CBV of 0.86) contains 13 statements and is very diverse as can be seen from the bridging value range from 0.66 to 1.00. The cluster describes aspects of acceptance of LA and their results by different stakeholders but also covers the comparability of methods and the dependence of LA on context and objectives. An example statement is “that administrators invest in scaling successful tools across their programming”.

Cluster 4 *Learning outcome* (CBV of 0.46) is also diverse, with a bridging value range from 0.19 to 0.87. It contains 13 statements that deal with comparability of LA results, teacher motivation, result accuracy and feedback for teachers, e.g., “if teachers are able to gain new insights using the given LA methods”, “that LA results are compared with other (traditional) measures”.

Cluster 5 *Teacher awareness* (CBV of 0.41) consists of twelve statements with bridging values from 0.18 to 0.73. Most statements are connected to teachers changing their course material or their teaching behaviour in response to LA results about their students: “that teachers change their behaviour in some aspects”, “that teachers react in a more personalized way to how their students are dealing with learning material”.

Cluster 6 *Learning performance* (CBV of 0.31) is one of the smallest clusters, as it consists of only eight statements. The bridging value range is relatively small, i.e., 0.11 to 0.59. Statements in this cluster are about student performance, learning and achievement improvement. Representative statements are “that change in workplace learning is measurable”, “the extent to which the achievement of learning objectives can be demonstrated”.

Cluster 7 *Learning support* (CBV of 0.45) is a very stable but also rather large cluster with 18 statements. Its bridging values range from 0.14 to 0.76. Statements in this cluster are often formulated generally and deal with support for teachers as well as for students, e.g., “an early detection of students at risk”, “the ability to explain what could help to further improve”, “that students regularly utilize the tools provided”.

Cluster 8 *Student awareness* (CBV of 0.11) is the largest and most coherent cluster. It contains 20 statements and its bridging value range is from 0.00 to 0.43. The cluster is also very stable and consistent. All statements are related to students, their achievement, success, self-regulation, awareness, learning behaviour and motivation, e.g., “that students become more self-regulated in their learning processes”, “that students are more aware of their learning progress”.

Rating Maps

Once the cluster map had been agreed, the experts’ ratings of the quality indicators could be included in the calculation. Experts were asked to rate *importance* and *feasibility* on a scale from one to seven (one for a low, seven for a high rating). The GCM tool automatically applies the experts’ ratings to the cluster map, indicating the importance or feasibility of each cluster by layering them. The system always divides the ratings into five layers based on the average ratings provided by the participants for the rating maps. The anchors for the map legend are based on the high and low average ratings across all of the participants. One layer indicates a low rating, whereas five layers indicate a high rating of the respective aspect.

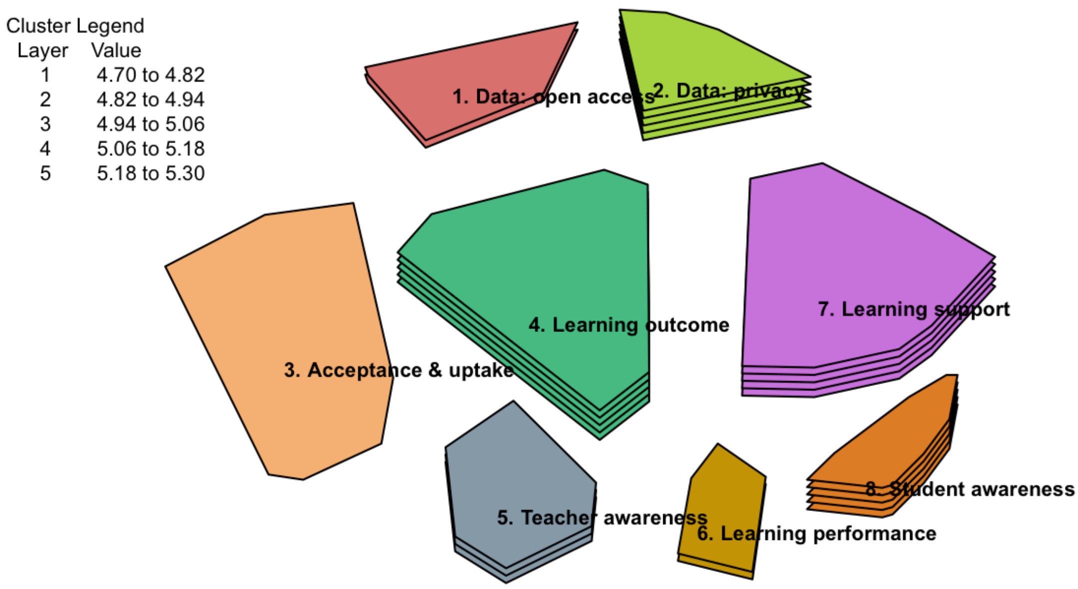

Figure 4: Rating map on importance

Figure 4 shows the rating map according to the *importance* aspect. Clusters *Data: privacy*, *Learning outcome*, *Learning support*, and *Student awareness* each received very high importance ratings as they all have five layers. *Teacher awareness* has three layers, while *Data: open access* and *Learning performance* have two layers each and *Acceptance & uptake* the least coherent cluster, has only one layer.

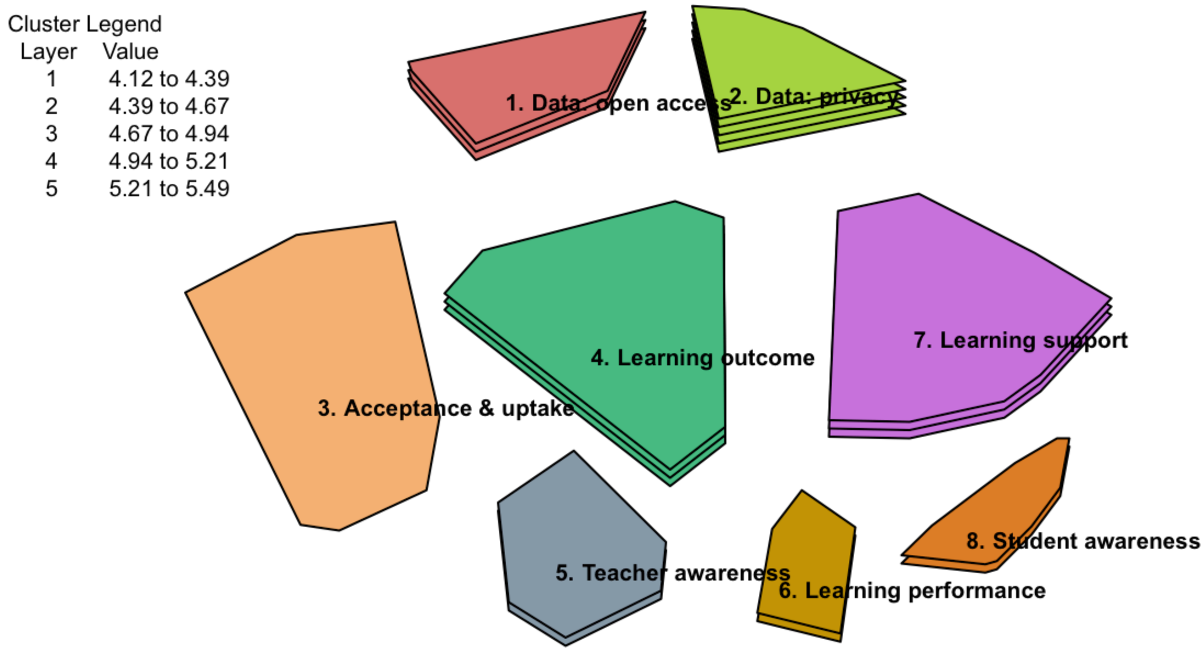

Figure 5: Rating map on feasibility

Looking at the *feasibility*-rating map (Figure 5) one can see a shift in the ratings assigned by the experts. Although the Data: privacy cluster is again assigned five layers because it is deemed highly feasible by the experts, the other three very important clusters have been rated less feasible: *Learning outcome* and *Learning support* only receive an intermediate level of feasibility with three layers each. *Student awareness*, a highly important cluster, receives a low feasibility rating with two layers only. *Teacher awareness* also drops down to two layers. The cluster dealing with *Acceptance & uptake* was seen as neither important nor feasible by the experts. The only cluster that receives more layers in the *feasibility*-rating map is *Data: open access*, which is thus deemed more feasible than it is important.

A ladder graph (see Figure 6) offers a form of visualisation that is well suited to comparison of the clusters' ratings according to *importance* and *feasibility*. The rating values are based on a cluster's average rating. A Pearson product-moment correlation coefficient (r = 0.65) indicates a strong positive relationship between the two aspects of importance and feasibility. For both ratings, the *Data: privacy* cluster receives the highest values while *Acceptance & uptake* receives the lowest. As was observable from the rating maps, the three clusters related to *Learning outcome*, *Learning support* and *Student awareness* have all been rated as important but as less feasible.

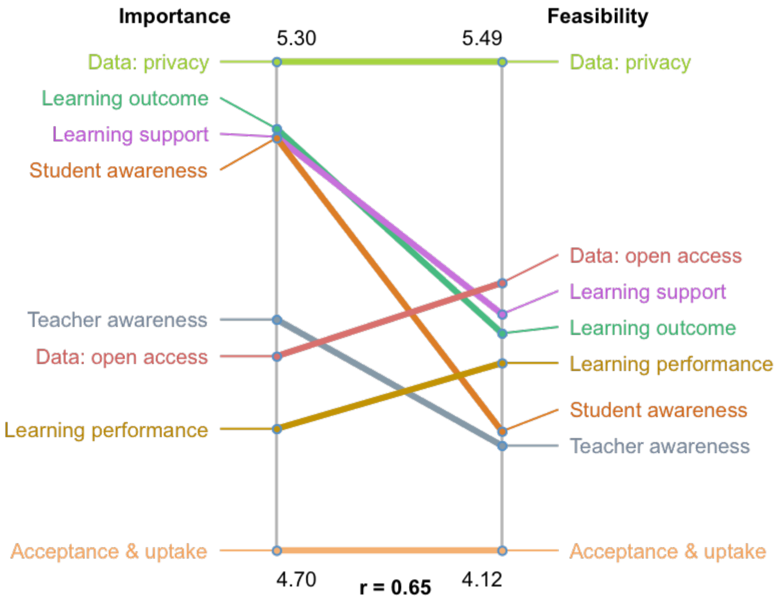

Figure 6: Ladder graph of the rating values for the clusters

A third visualisation the GCM tool presents go-zones. These are bivariate graphs that support deeper exploration of the statements in relation to their ratings. A Go-zone graph uses the mean values of the ratings for importance and feasibility to map each statement onto a space in which the x axis represents *importance* and the y axis represents *feasibility*. Go-zone graphs were created for all statements and for individual clusters. Figure 7 shows the go-zone graph for all 103 statements. Go-zone graphs are useful in the selection of suitable quality indicators because they highlight those statements with a good balance of *importance* and *feasibility*. These are the statements that appear in the upper right quartile of the graph. When deciding on quality indicators for the individual criteria, a sensible approach involves choosing statements from the mainly feasible or mainly important quadrant only if they are close enough to the upper right quadrant and support respective criterion.

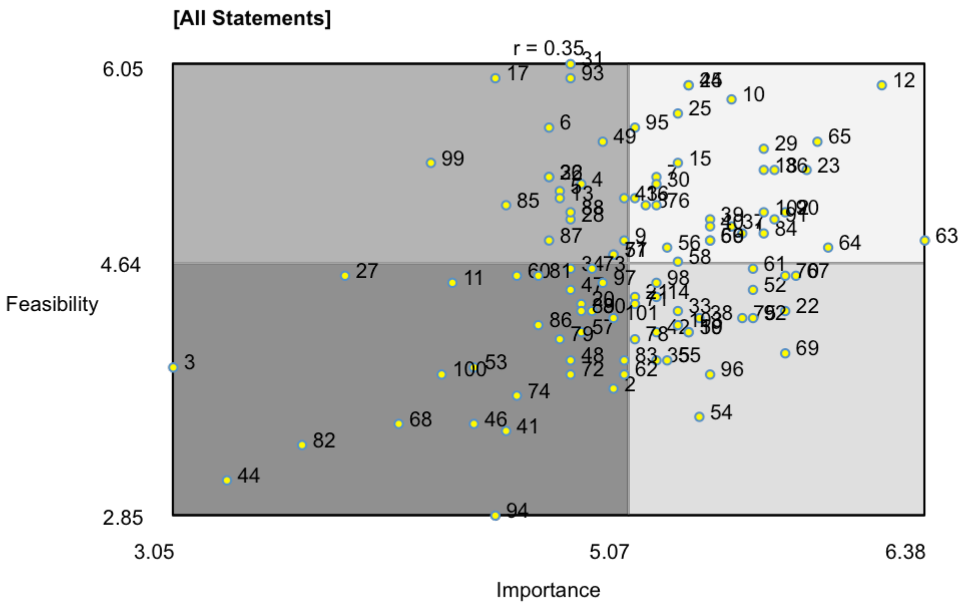

Figure 7: Go-zone graph of all 103 statements

The Framework

Discussion of the GCM Results

Looking at the clusters in Figure 3, their coherence is clear. The four most coherent clusters (*Data: open access*, *Data: privacy*, *Learning performance* and *Student awareness*) are the ones with the smaller bridging values. These clusters are therefore smaller in area than the others. The three least coherent clusters (*Acceptance & uptake*, *Learning outcome* and *Learning support*), are the ones with higher bridging values, and so their area is much larger. Two clusters, *Learning support* and *Student awareness*, were the only two to remain stable until the five-cluster solution was tested. This stability implies close agreement between the experts' sorting and the system's multidimensional scaling and hierarchical clustering. We take cluster coherence and stability to be a first indication of relevance when trying to find quality indicators for learning analytics.

Comparing the two rating maps (see Figure 4 and Figure 5) yields other insights. The *Acceptance & uptake* cluster received low ratings for *importance* as well as for *feasibility*. The experts' low rating is also supported by the cluster's coherence. With an average bridging value of 0.86 and individual statement bridging values from 0.66 up to 1.00, the cluster contains a diverse collection of statements. When identifying quality indicators to evaluate effects of learning analytics, we therefore focused on all other clusters first in order to find suitable criteria before taking this cluster into account as the indicators it contains are too incoherent, too vague, unimportant and unfeasible as a group.

The comparison of the maps leaves us with a slightly different cluster landscape: The two clusters in the North (1, 2) both deal with data, access, methods, algorithms, transparency and privacy, i.e., with technical issues, while the clusters in the South (5, 6, 8) deal with awareness, reflection, performance and behavioural change of students and teachers, i.e., with human issues. The "technical North" (*Data: open access* and *Data: privacy*) and the "human South" (*Teacher awareness*, *Learning performance* and *Student awareness*) are bridged by a wide layer of learning-related clusters (*Learning outcome* and *Learning support*). Apart from the North-South view, one can also look at the map with an East-West perspective: The three Eastern clusters (*Data: privacy*, *Learning support* and *Student awareness*) are more concerned with issues during the learning process while the Western clusters (*Data: open access*, *Learning outcomes*, *Teacher awareness* and *Learning performance*) are slightly more concerned with issues of learning output and results. These groupings show a thematic tendency. While constructing the framework, we concluded that the aspects of technology, stakeholders (humans), learning processes and learning outcomes should all be reflected in the criteria.

Taking the two rating aspects *importance* and *feasibility* into account, we get two different versions of the landscape described above. The *importance* map on the left side of Figure 8 clearly shows that the learning-related middle layer – the clusters related to *Learning outcome* and *Learning support* that are shown within the dashed line is deemed highly important by the LA experts. But all Eastern clusters– the ones related to *Data: privacy*, *Learning support* and *Student awareness* that are shown within the dotted line – received five layers of *importance*. Generally, the focus of *importance* is on the learning process-related clusters.

For the *feasibility* map on the right side of Figure 8, the landscape shifts. Now there is a clear North-South divide: The technically-oriented clusters in the North (dotted oval) are deemed most feasible by the experts, followed by the learning-related layer in the middle (short-dashed oval) and then by the human-related clusters in the South (long dashed circle). This again supports the construction of criteria for the framework according to the data-learning support and process-stakeholder view.

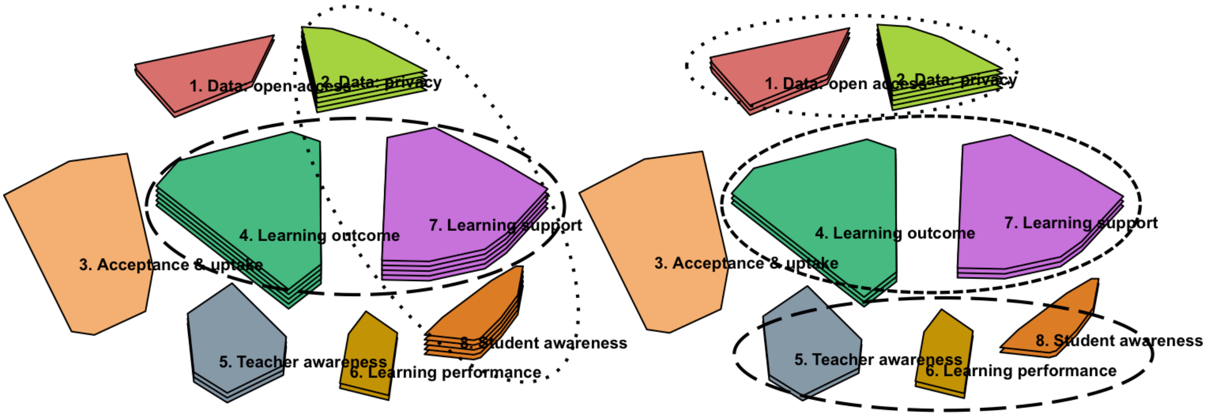

Figure 8: Comparison of the rating maps of importance (left) and feasibility (right)

Looking at the ladder graph in Figure 6 allows a closer look at the differences in average ratings for the different clusters. The lower rating for *feasibility* than for *importance* for a number of clusters is clear. The most striking difference in ratings applies to the *Student awareness* cluster. The experts consider that it is fairly important to take student awareness into account when evaluating LA, but consider that it is difficult to apply in practice. This may be because many teachers do not consider students to be accurate judges of their own learning processes and progress (Drachsler & Greller, 2012).

When deciding the criteria of the framework it was important to balance *importance* ratings and *feasibility* ratings. Due to the relatively high *importance* of the clusters about *Data: privacy*, *Learning outcome*, *Learning support* and *Student awareness*, they were used as the basis for the criteria of the framework. The *feasibility* ratings of the clusters were then be used to associate the remaining clusters with these four criteria: The *Data: privacy* cluster is considered the most feasible, followed by the *Data: open access* cluster. The two were therefore combined to form one data criterion. The next two clusters on the *feasibility* scale are *Learning support* and *Learning outcome*. These remained separate, due to their high *importance* rating. As the latter cluster is followed by the *Learning performance* cluster and as they both deal with learning results and effects, it made sense to use them to construct a combined criterion. The next cluster on the *feasibility* scale is *Student awareness*, closely followed by *Teacher awareness*. Both of these are “human clusters”, concerned with awareness, reflection and behavioural change. They were therefore combined into one criterion, even though they address different stakeholders.

Outline of the Framework

The results of the GCM study were used to identify four topic areas that could be turned into criteria for the framework.

The first topic area, *Data Aspects*, deals with anything related to data, algorithms, transparency and privacy. It is based on the clusters *Data: privacy* and *Data: open access*. It contains the quality indicators *Transparency*, *Data Standards*, *Data Ownership*, and *Privacy*.

The second topic area, *Learning Support*, is entirely based on the cluster of that name. It is concerned with support for students and teachers while using LA tools during the learning process. The quality indicators of this criterion are *Perceived Usefulness*, *Recommendation*, *Activity Classification*, and *Detection of Students at Risk*.

The third topic area, *Learning Measures and Output*, deals with the results of the learning process, including any issues of output, consequence, performance or outcome. This topic area does not relate primarily to individual student performance and grades, but instead refers to the LA tools' results and outcomes. It is made up of the two clusters *Learning outcome* and *Learning performance*, and contains the quality indicators *Comparability*, *Effectiveness*, *Efficiency*, and *Helpfulness*.

The fourth topic area, *Objectives*, is concerned with the educational aims identified at the beginning of this report. It contains the quality indicators *Awareness*, *Reflection*, *Motivation*, and *Behavioural Change* of students and educators during the learning process.

Most statements related to stakeholders that were used during the first stage of topic generation were about learners and teachers. Very few related to institutions. This was because the disparate e cluster *Acceptance & uptake*, which contained several statements about institutions, was not originally taken into account. It was also because statements relating to institutions were not grouped together, but were spread cross clusters. As we consider indicators of organisational issues to be an important aspect when considering the evaluation of LA tools (cf. Arnold et al., 2014), we added a fifth topic area, *Organisational Aspects,* to the framework. This contained the quality indicators *Availability*, *Implementation*, *Training of Educational Stakeholders*, and *Organisational Change*.

These quality indicators are based on a review of the statements in the go-zone graphs of each cluster. In most cases, these indicators are situated in the upper right quadrant of the go-zone graphs. In some cases, statements from the *feasible* or *important* quadrants were also chosen, if they were located close to the quadrant containing indicators that are both *important* and *feasible*. The statements chosen for each criterion were then combined and rewritten as slightly shorter, more general statements that clearly represent a quality indicator for a given criterion. Figure 9 shows a first outline of the five criteria, each with their four quality indicators.

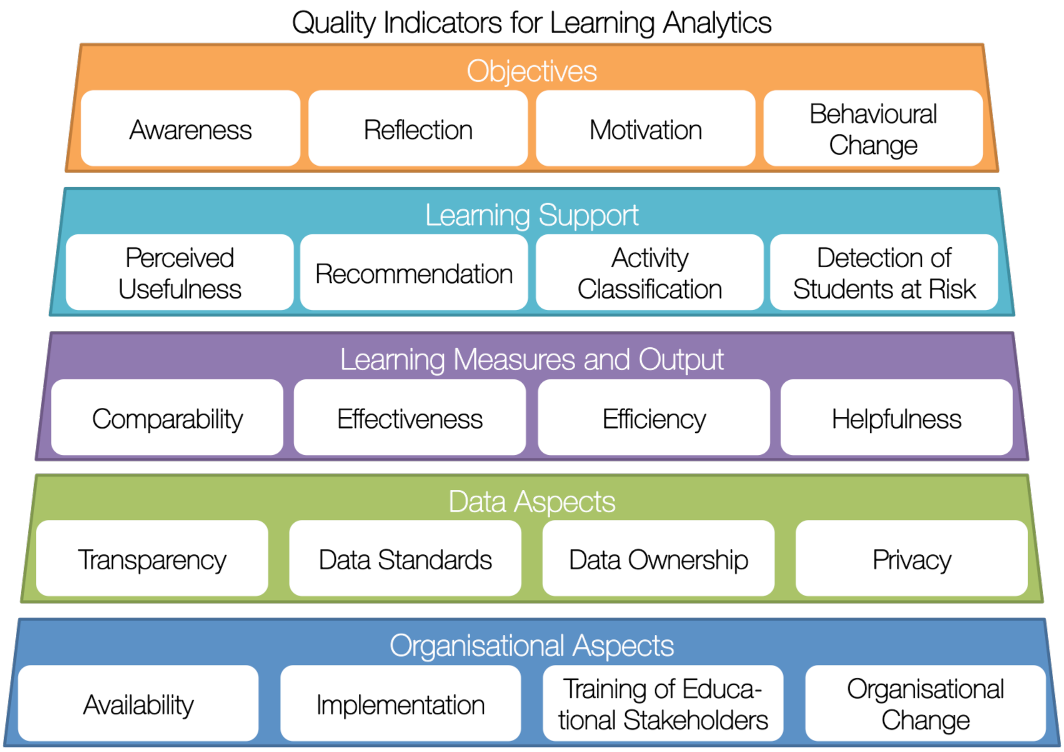

Figure 9: First outline of the framework of quality indicators for learning analytics

Turning the Framework into an applicable Tool

Evaluation Study Methodology

For the evaluation of the framework two things had to be done: on the one hand the framework needed to be turned into an applicable tool itself and on the other hand a collection of learning analytics tools to validate the framework against had to be compiled. As a first step, the framework's criteria and quality indicators were therefore transformed into a questionnaire. Every criterion in the framework was given its own section with questions corresponding to the individual quality indicators. For every quality indicator we asked whether it was supported by the tool in question or not or whether it was not applicable to the tool, e.g. whether a tool detected students at risk or not. In case a quality indicator was supported by the tool, it was also asked in what way that support was handled, e.g. that a tool fosters awareness for students and teachers as its analyses are available to both.

As the main aim of this study was not to actually compare existing learning analytics tools with one another but to use the comparison to evaluate the framework, a rating question for each quality indicator was added as well, asking participants to rate how difficult or easy it was to judge/apply a quality indicator on a scale from 1 (very difficult) to 5 (very easy). At the end of each criterion section participants were offered an open text box asking for any additional comments about any of the quality indicators of that section or issues they felt had not been addressed by the evaluation questions. The complete questionnaire can be accessed at *http://bit.ly/EFqiLA*.

To find suitable learning analytics tool candidates the submissions to the Learning Analytics and Knowledge conferences as well as a number of existing tools were browsed. In the end, eight prominent learning analytics tools were randomly selected to be used for the evaluation of the framework: Blackboard Learn 9.1 Retention Centre (Blackboard, 2014), CourseSignals[1] (Arnold, 2010; Arnold & Pistilli, 2012), EnquiryBlogger[2] (Buckingham et al., 2012; Ferguson et al., 2011), the LeMo project[3] (Elkina et al., 2013; LeMo, n.d.), SNAPP[4] (Baron & Jayaprakash, 2014), StepUp! (Santos et al., 2012; Santos et al., 2013), Student Activity Meter (Govaerts et al., 2011; Govaerts et al., 2012) and Student Explorer (Aguilar et al, 2014; Lonn & Teasley, 2014). The study was conducted with members from the LACE project consortium and its associated partners. Each of the eight participants evaluated two of the eight tools, which in turn meant that each of the eight tools was evaluated twice.

Due to the nature of the study, i.e. turning the framework into a tool and evaluating it, analysis outcomes dealing with the eight individual tools will not be addressed in this deliverable. Instead the focus is set entirely on the setup and applicability of the framework's criteria and quality indicators.

Results

To get an overview of the questionnaire results for all quality indicators Figure 10 shows how many *yes*, *no* and *not applicable* every quality indicator received while the rating values of all quality indicators and their average rating are listed in Figure 11. The highest average rating was achieved

---

[1] *http://www.itap.purdue.edu/learning/tools/signals/* and *http://www.itap.purdue.edu/studio/signals/*
[2] *http://learningemergence.net/tools/enquiryblogger/*
[3] *http://www.lemo-projekt.de*
[4] *http://www.snappvis.org*

by the quality indicator *awareness*, i.e. 4.3, while the lowest average was achieved by *efficiency*, i.e. 2.6. These two indicators were also the ones with the lowest (*awareness*) and highest (*efficiency*) non-applicability.

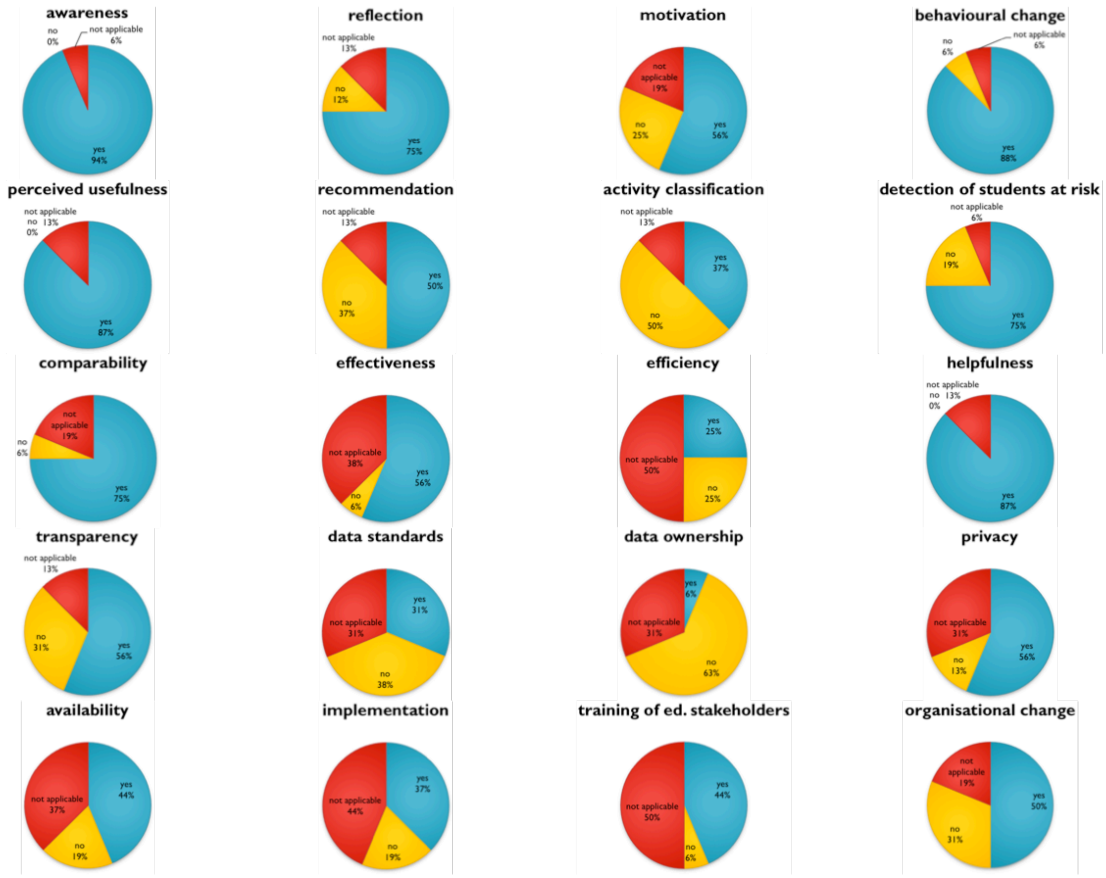

Figure 10: Visualisation of the presence of quality indicators in a tool ordered according to the QI's position in Figure 4

	1	2	3	4	5	avg.
awareness	-	1	1	7	7	***4.3***
reflection	1	-	2	6	7	4.1
motivation	1	3	3	4	5	3.6
behavioural change	-	3	5	5	3	3.5
perceived usefulness	2	-	1	7	6	3.9
recommendation	1	1	3	4	7	3.9
activity classification	4	3	3	1	5	3.0
det. of students at risk	-	1	3	6	6	4.1
comparability	-	6	2	5	3	3.3
effectiveness	2	5	4	4	1	2.8
efficiency	4	3	5	3	1	***2.6***
helpfulness	2	4	2	5	3	3.2
transparency	-	4	6	5	1	3.2
data standards	3	2	2	5	4	3.3
data ownership	3	3	3	6	1	2.9
privacy	-	3	2	8	3	3.7
availability	2	1	1	3	9	4.0
implementation	2	1	2	2	9	3.9
training of stakeholders	2	-	1	8	5	3.9
organisational change	2	-	1	12	1	3.6

Figure 11: Overview of 1-to-5 scale ratings plus average ratings for all QIs

The results of the framework evaluation study allow us to identify several issues with the framework that need to be addressed in order to work towards an improved evaluation framework. Some of them might be cleared up fairly easily while others will need to be examined carefully so as to ensure actual improvement. The issues identified can be divided into the following categories: (1) *concept definitions*, (2) *differentiations*, (3) *framework structure*, and (4) *questionnaire adaption*.

The first category, *concept definitions*, relates to any case where the study participants have expressed that either a criterion or a quality indicator needs to be rephrased or defined more clearly as they currently do not convey enough information about themselves in order to be properly applied to a tool evaluation. Participants particularly mentioned one criterion and three quality indicators where this is the case: *Learning Measures and Output*, *activity classification*, *comparability*, and *data ownership*. Renaming, and thus redefining, a whole criterion also influences how the quality indicators of that criterion are interpreted. When constructing the next version of the applicable framework, this will have to be taken into account. Although only these four elements of the framework have explicitly been mentioned, all other criteria and quality indicators will also undergo a careful definition inspection and pilot testing to avoid similar issues with the next framework version.

The issues of the second category, *differentiations*, are closely related to those of the first category. Participants identified some quality indicators, or better pairs of quality indicators, that need to be defined more clearly and to be supported by some distinct example so as to be able to properly distinguish between them. Otherwise users of the framework tool might misunderstand them and thus distort the results of a LA tool evaluation. The quality indicators mentioned by the study participants are *usefulness* vs. *helpfulness*, *effectiveness* vs. *efficiency*, *transparency* vs. *privacy*, and *availability* vs. *implementation*.

The third category, *framework structure*, deals with the issue of inter-criterion homogeneity of quality indicator types. Participants of the evaluation study suggested ensuring that the types of indicators within one criterion are the same in order to improve the applicability of the whole criterion. Generally, indicators should tend to be concept rather than feature driven. Participants identified this issue in the criterion *Learning Support* where three indicators, i.e. *recommendation*, *activity classification* and *detection of students at risk*, can be termed as features or functionalities of a tool while the fourth indicator, i.e. *perceived usefulness*, can be termed as a goal of a tool. Although participants identified this issue only in this criterion, all other criteria will be inspected as well so as to avoid this issue from appearing again in the next evaluation cycle.

The fourth category, *questionnaire adaption*, comprises issues that need to be addressed when setting up the next version of the framework's questionnaire or better the next practical, applicable and executable version of the framework (as this might not be the same questionnaire format as it was this time). Participants of the evaluation study noted several aspects that would highly improve the applicability of the indicators. For many indicators the answers would differ depending on the user type addressed, i.e. answers for learners, educators or administrative staff can vary. This should thus be set clear when starting a questionnaire and might lead to specific instances of the evaluation framework being needed for different stakeholders. When formulating questions for the indicators, these should best ask about the intention of a tool as this is something that can be answered much more easily than a tool's actual impact on a user. This is especially true if the evaluator has no access to the tool but has to work with descriptive resources. The third issue related to questionnaire adaption is the possible addition of answer options. Several participants of the evaluation study remarked that they would have liked to see an *I don't know*- option or a *too much*-option as some indicators sometimes can neither be marked as *not there* nor as *not applicable* due to too sparse information.

An issue that is not related to any of the categories and that cannot be improved by us is the sparsity of information provided in the resources about learning analytics tools. While addressing the issues mentioned above will make it easier for externals to evaluate a LA tool, the most complete evaluations will be those of the actual users or creators of a tool. In those cases where users or creators apply the framework to their own tool, however, the results might be biased which has to be taken into account as well.

Conclusion

The results of a group concept mapping study – including rating maps, ladder graph and go-zone graphs – have been used to create a framework for the evaluation of learning analytic tools (see Figure 9). This framework has five criteria, each with four quality indicators. These are:

1. *Objectives* (*Awareness*, *Reflection*, *Motivation*, *Behavioural Change*)
2. *Learning Support* (*Perceived Usefulness*, *Recommendation*, *Activity Classification*, *Detection of Students at Risk*)
3. *Learning Measures and Output* (*Comparability*, *Effectiveness*, *Efficiency*, *Helpfulness*)
4. *Data Aspects* (*Transparency*, *Data Standards*, *Data Ownership*, *Privacy*)
5. *Organisational Aspects* (*Availability*, *Implementation*, *Training of Educational Stakeholders*, *Organisational Change*).

This initial version of the framework was turned into an evaluation tool that was then evaluated and tested by the LACE project consortium for its applicability. In order to do so, we created a questionnaire based on the framework (available at *bit.ly/EFqiLA*) that was circulated among project members as well as Associated Partners of the LACE project. With the feedback from the participants we were able to identify problematic issues and have collected suggestions on how to overcome the issues and improve the framework.

The outcomes of the evaluation study will be carefully analysed and discussed within the LACE consortium to develop an improved version of the framework. Apart from the theoretical framework set up, the structure of the related evaluation instrument will also be improved as different stakeholders might require different versions of the instrument. The improved framework as well as its implementation will then form the basis of another evaluation cycle. The resulting updates to the framework will be reported in 2015.

The results of the tool analyses of these studies will be fed into the LACE Evidence Hub. This knowledge base of evidence is created and curated by the LACE project and captures evidence for the effectiveness and the relative desirability of the outcomes resulting from the use of various tools and techniques.

References

Aguilar, S., Lonn, S., & Teasley, S. D. (2014). Perceptions and use of an early warning system during a higher education transition program. In K. E. Arnold, S. Teasley & A. Pardo (Eds.), *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (LAK'14)* (pp. 113–117). New York, NY: ACM.

Arnold, K. E. (2010). Signals: Applying academic analytics. *EDUCAUSE Quarterly, 33*(1).

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at purdue: Using learning analytics to increase student success. In S. Buckingham Shum, D. Gasevic & R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12)* (pp. 267–270, New York, NY: ACM.

Arnold, K. E., Lonn, S. & Pistilli, M. D. (2014). An exercise in institutional reflection: The learning analytics readiness instrument (LARI). In K. E. Arnold, S. Teasley & A. Pardo (Eds.), *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (LAK'14)* (pp. 163–167). New York, NY: ACM.

Baron, J., & Jayaprakash, S. (2014). Snapp (social network analysis & pedagogical practices) for sakai cle v2.8x & v2.9x. Retrieved October 29, 2014 from https://confluence.sakaiproject.org/x/MYEPBQ

Blackboard (2014). *Student performance and retention*. Retrieved October 29, 2014 from https://help.blackboard.com/en-us/Learn/9.1_2014_04/Instructor/130_Student_Performance

Bolton, G. (2010). *Reflective practice: Writing & professional development* (3rd ed.). London, UK: Sage.

Buckhingham Shum, S., Ferguson, R., & Deakin Crick, R. (2012). *Enquiryblogger: Blog-based learning analytics for learning power & authentic enquiry*. Retrieved October 29, 2014 from http://de.slideshare.net/sbs/ enquirybloggeranalyticscalrg2012

Concept Systems Global. (n.d.). *Concepts systems global*. Retrieved October 29, 2014, from http://www.conceptsystemsglobal.com

Drachsler, H., & Greller, W. (2012). The pulse of learning analytics: Understandings and expectations from the stakeholders. In S. Buckingham Shum, D. Gasevic & R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12)* (pp. 120–129) New York, NY: ACM.

Drachsler, H., Stoyanov, S., d'Aquin, M., Herder, E., Guy, M., & Dietze, S. (2014). An Evaluation Framework for Data Competitions in TEL. In Ch. Rensing, S. de Freitas, T. Ley and P. J. Muñoz-Merino (Eds). *Proceedings of EC-TEL 2014* (LNCS) (Vol. 8719). Heidelberg, Germany: Springer.

Elkina, M., Fortenbacher, A., & Merceron, A. (2013). The learning analytics application lemo - rationals and first results. *International Journal of Computing, 12*(3), 226–234.

Endsley, M. R. (1995). Towards a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32–64. doi:10.1518/001872095779049543

Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 3–29). Mahwah, NJ: Lawrence Erlbaum Associates.

Ferguson, R., Buckhingham Shum, S., & Deakin Crick, R. (2011). Discussion paper: Enquiryblogger - using widgets to support awareness and reflection in a ple setting. In Workshop on Awareness and Reflection in Personal Learning Environments, PLE Conference 2011.

Ferguson, R. (2014, March 26). Learning analytics don't just measure students' progress – They can shape it. *theguardian.com*, Retrieved October 29, 2014, from http://www.theguardian.com/education/2014/mar/26/learning-analytics-student-progress

Govaerts, S., Verbert, K., & Duval, E. (2011). Evaluating the student activity meter: Two case studies. In H. Leung, E. Popescu, Y. Cao, R. Lau, and W. Nejdl (Eds.), *Advances in Web-Based Learning - ICWL 2011* (LNCS) (Vol. 7048, pp. 188–197). Heidelberg, Germany: Springer.

Govaerts, S., Verbert, K., Duval, E., & Pardo, A. (2012). The student activity meter for awareness and self-re ection. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems, CHI EA '12*, pages 869–884, New York, NY: ACM.

Kane, M., & Trochim, W. M. K. (2007). *Concept mapping for planning and evaluation*. Thousand Oaks, CA: Sage Publication.

LeMo (n.d.). Lemo handbuch fuer anwender und administratoren. Retrieved October 29, 2014 from http://lemo.htw-berlin.de/public/doc/LeMo_HB_final_a.pdf

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46*(5), 31–40. Retrieved August 15, 2014, from https://net.educause.edu/ir/library/pdf/ERM1151.pdf

Lonn, S., & Teasley, S. D. (2014). Student explorer: A tool for supporting academic advising at scale. In *Proceedings of the First ACM Conference on Learning @ Scale Conference, L@S '14*, pages 175–176, New York, NY: ACM.

Santos, J. L., Verbert, K., & Duval, E. (2012). Empowering students to reflect on their activity with stepup!: two case studies with engineering students. In *Proceedings of the 2nd workshop on Awareness and Reflection, ARTEL '12*, Saarbrücken, Germany, 2012. CEUR.

Santos, J. L., Verbert, K., Govaerts, S., & Duval, E. (2013). Addressing learner issues with stepup!: An evaluation. In D. Suthers, K. Verbert, E. Duval & X. Ochoa (Eds.), *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13)* (pp. 14–22), New York, NY: ACM.

Scheffel, M., Niemann, K., Pardo, A., Leony, D., Friedrich, M., Schmidt, K., … Delgado Kloos, C. (2011). Usage pattern recognition in student activities. In C. Delgado Kloos, D. Gillet, R. M. Garcia Crespo, F., Wild & M. Wolpers (Eds.), *Proceedings of EC-TEL 2011*(LNCS) (Vol. 6964, pp. 341–355). Heidelberg, Germany: Springer.

Scheffel, M., Niemann, K., Leony, D., Pardo, A., Schmitz, H.-C., Wolpers, M., & Delgado Kloos, C. (2012). Key action extraction for learning analytics. In A. Ravenscroft, S. Lindstaedt, C. Delgado Kloos & D. Hernandez-Leo (Eds.), *Proceedings of EC-TEL 2012* (LNCS) (Vol. 7563, pp. 320–333). Heidelberg, Germany: Springer.

Scheffel, M., Drachsler, H., Stoyanov S., & Specht, M. (2014). Quality indicators for learning analytics. *Educational Technology & Society, 17*(4), 117–132.

Schön, D. (1983). *The reflective practitioner: How professionals think in action*. London, UK: Temple Smith.

Zinn, C., & Scheuer, O. (2006). Getting to know your student in distance learning contexts. In W. Nejdl & K. Tochtermann (Eds.), *Proceedings of EC-TEL 2006* (LNCS) (Vol. 4227, pp. 437–451). Heidelberg, Germany: Springer.

About

Version History

Date	Notes	Person
2014-09-09	Basic document structure set up in the template. Incorporated initial sections.	Maren Scheffel, OUNL
2014-09-30	Additional text and images	Maren Scheffel, OUNL
2014-10-02	Review comments	Tore Hoel, HIOA
2014-10-02	Review comments	Rebecca Ferguson, OUUK
2014-10-29	Finalisation	Maren Scheffel, OUNL
2014-10-31	Cleared for submission to the EC Project Office	Hendrik Drachsler, OUNL

About this document

Some research reported in this deliverable has been published in (Scheffel et al., 2014) while others have been submitted to the research track of the Learning Analytics & Knowledge (LAK) Conference 2015.

For more information, see the LACE Publication Policy *http://www.laceproject.eu/publication-policy/.*

About LACE

The LACE project brings together existing key European players in the fields of learning analytics & educational data mining who are committed to building communities of practice and sharing emerging best practice in order to make progress towards four objectives.

> *Objective 1 – Promote knowledge creation and exchange*
> *Objective 2 – Increase the evidence base*
> *Objective 3 – Contribute to the definition of future directions*
> *Objective 4 – Build consensus on interoperability and data sharing*

http://www.laceproject.eu *@laceproject*

This document was produced with funding from the European Commission Seventh Framework Programme as part of the LACE Project, grant number 619424.