Widget, widget as you lead, I am performing well indeed! Using results from an exploratory offline study to inform an empirical online study about a learning analytics widget in a collaborative learning environment

Maren Scheffel
Open Universiteit
Valkenburgerweg 177
6419 AT Heerlen, NL
maren.scheffel@ou.nl

Hendrik Drachsler
Open Universiteit
Valkenburgerweg 177
6419 AT Heerlen, NL
hendrik.drachsler@ou.nl

Karel Kreijns
Open Universiteit
Valkenburgerweg 177
6419 AT Heerlen, NL
karel.kreijns@ou.nl

Joop de Kraker
Open Universiteit
Valkenburgerweg 177
6419 AT Heerlen, NL
joop.dekraker@ou.nl

Marcus Specht
Open Universiteit
Valkenburgerweg 177
6419 AT Heerlen, NL
marcus.specht@ou.nl

ABSTRACT

The collaborative learning processes of students in online learning environments can be supported by providing learning analytics-based visualisations that foster awareness and reflection about an individual's as well as the team's behaviour and their learning and collaboration processes. For this empirical study we implemented an activity widget into the online learning environment of a live five-months Master course and investigated the predictive power of the widget indicators towards the students' grades and compared the results to those from an exploratory study with data collected in previous runs of the same course where the widget had not been in use. Together with information gathered from a quantitative as well as a qualitative evaluation of the activity widget during the course, the findings of this current study show that there are indeed predictive relations between the widget indicators and the grades, especially those regarding responsiveness, and indicate that some of the observed differences in the last run could be attributed to the implemented activity widget.

CCS Concepts

•Applied computing → Collaborative learning; E-learning; •Human-centered computing → User studies; Visualization systems and tools; •General and reference → Evaluation;

LAK '17, March 13 - 17, 2017, Vancouver, BC, Canada

ISBN 978-1-4503-4870-6/17/03...$15.00

DOI: http://dx.doi.org/10.1145/3027385.3027428

Keywords

learning analytics; statistical analysis; tool evaluation

1. INTRODUCTION

One way to support the collaborative learning processes of student teams in virtual learning environments is to provide explicit information to the students about the activities of the group members and to stimulate awareness, reflection and social interaction [11]. Although using behavioural data automatically collected from the learning environment is not to be seen as a one-to-one replacement for using subjective data collected via questionnaires or interviews [8], making use of learning analytics based on interaction data does have the advantage of being non-disruptive and covering the whole student population of a course. A learning analytics widget in a computer-supported collaborative learning environment can thus provide feedback [9, 14] to students as well as teachers by visualising the students' activities within the virtual learning environment in order to facilitate awareness and reflection[15].

Endsley [5, 6] describes being aware of one's own situation as a three level process: (i) perceiving the elements in the current situation, (ii) comprehending the situation, and (iii) projecting what a future status could look like. Once awareness of the situation is established, a user can reflect on it in relation to his behaviour [19] and can subsequently adapt or even change his behaviour if necessary. According to McAlpine & Weston [13] reflection is to be seen as a mechanism that can improve teaching and thus maximise learning and not as an end in itself. Reflection processes and behavioural change are, however, not only influenced by awareness [2]. Whenever someone engages in self-regulated learning, they bring their own knowledge, beliefs and skills into the process [23]. Additionally, emotions, the social environment as well as one's own behaviour play a role [25]. The way in which someone acts and reacts in a given situation thus depends on the different ways they have constructed their current knowledge [24].

The relevance of these aspects has been emphasised by Verbert et al. [22] in their process model for learning analytics applications that consists of four stages: awareness, reflection, sensemaking, and impact. As the discussion about the effect of learning analytics and the need for empirical studies has increased [20, 7], a number of recent studies have investigated the impact of learning analytics dashboards on different aspects, e.g. individual goal attainment and motivation. Lonn et al. [12] investigated whether the motivation of students in a summer bridge program, i.e. students among the at-risk population in postsecondary education, was affected by the use of learning analytics. Their findings suggest that being exposed to a learning analytics application displaying their academic performance can negatively predict the change of mastery orientation, i.e. it decreases, and can thus affect a student's interpretation of their data and their success. The authors stress that student goal perception and formative performance thus need to be carefully considered when designing learning analytics interventions.

Beheshitha et al. [1] also examined the effect of learning analytics visualisations. Their experiment took place in a blended course setting where each student was randomly assigned to one of three available visualisations. The results revealed that the visualisations had different, i.e. positive or negative, effects on the quality and quantity of forum posts by the students that depended on the students' achievement goal orientation. These authors stress that it is important to consider individual differences such as achievement goal orientation in the design process of learning analytics visualisations. A third study by Khan and Pardo [10] showed that students use learning analytics differently, i.e. depending on their information need or the learning activity or phase. All three of these studies clearly emphasise that for learning analytics visualisations to have a positive effect, they need to be embedded into the instructional design and that the students' personal preferences, e.g. goal attainment or motivation, need to be considered.

In order to add further results to the collection of empirical data studies, we have designed a learning analytics widget called 'activity widget' and implemented it into the learning environment of the European Virtual Seminar (EVS), an online course where geographically dispersed students work together on different topics in small teams. Based on data automatically collected in the EVS platform, the activity widget is made up of several radar and bar charts. The aim is to make students aware of the platform activity of their team in relation to their own activity level. Apart from making students aware, the activity widget also aims to foster reflection about how their behaviour can influence their position in the team and their course outcome.

1.1 Exploratory Offline Study

In a previous exploratory data study [17] (referred to as 'exploratory study' throughout this paper) we investigated the predictive power of several indicators of the activity widget towards the students' grades by instantiating these indicators with data from the four previous runs (2011-2012, 2012-2013, 2013-2014 and 2014-2015) of the European Virtual Seminar on Sustainable Development (EVS). That is, although the activity widget had not been used in those years, we analysed the log data from these years to explore what the widget indicator scores would have been if the widget had been used in those years. We tested whether the students' activity scores of the previous runs correlated with the tutor gradings and whether they validly reflected them. We did so for the whole run of the courses as well as for individual months.

More specifically, in the exploratory study we wanted to know (1) whether the widget indicator scores correlated with the tutor gradings of individual students at all, (2) whether the scores of some widget indicators were better predictors for the students' individual grades and (3) whether certain points in time produced indicator scores that are better grade predictors than others. We hypothesised that significant positive correlations existed between the widget indicators and the grades, that the widget indicator 'presence' (see explanation below) was a better predictor than the other ones and that the widget indicator scores produced in the second half of the course were better predictors towards the grades than those in the first half of the course.

The results of the correlation analysis and the structural equation modelling (SEM) of the exploratory study showed that most of the indicators indeed significantly and positively correlated with the grades and that they can be used as predictors. The scores of the 'presence' indicator, however, did not turn out to be better predictors for the grades, neither for the whole run nor for the individual months. Instead, the 'responsiveness' indicator achieved the best results. Looking at the individual months, the analysis showed that the months in the first half of the course yielded better correlation and SEM results than those in the second half. This unexpected outcome was due to an unforeseen large usage of communication tools outside of the course's learning environment. For detailed results and their discussion please refer to [17].

1.2 Approach

Keeping these results in mind, we implemented the activity widget into the learning environment of EVS and made it available to students and tutors in the 2015-2016 run of the course. In this current study (referred to as 'online study' throughout this paper) we investigated whether using the activity widget live in a run of the course yielded similar or different correlations between the widget indicator scores and the grades and whether the regression analyses performed in SEM showed approximately the same path-coefficients when compared to the exploratory study. The same set of analyses as used in the exploratory study was therefore applied to the data from the 2015-2016 run. The research questions that guided the correlation and regression analyses in our online study are:

RQ-A1: With the activity widget in use, do widget indicator scores again correlate significantly and positively with the tutors' gradings of individual students?

RQ-A2: With the activity widget in use, are the scores of the responsiveness indicator again better predictors for the students' individual grades than those of the others?

RQ-A3: With the activity widget in use, are the widget indicator scores produced in the first half of the course again better predictors than those produced in the second half?

As the activity widget aims at making students aware of their own activities relative to those of their fellow students as well as fostering reflection about how their behaviour influences their position within the team and the team's collaboration processes, we were interested in the users' experience with the widget during the 2015-2016 run. We therefore evaluated the activity widget using the Evaluation Framework for Learning Analytics (EFLA) questionnaire twice: the first evaluation was conducted in the middle of the course and the second one at the end. Using EFLA allowed us to take the students' as well as the tutors' points of view into account and to compare the two user groups with one another. The research questions that guided the widget evaluation are:

RQ-B1: Is there a difference in widget evaluation results between the mid-course questionnaire and the end-course questionnaire?

RQ-B2: Is there a difference in widget evaluation results between students and tutors?

The next section describes the course, the activity widget and the evaluation questionnaire in more detail and also elaborates on the method of analysis. After that, we present the results of our online study followed by a discussion and the conclusions.

2. METHOD

2.1 Participants and Materials

2.1.1 The EVS Course

Coordinated by the Open University of the Netherlands, the European Virtual Seminar on Sustainable Development (EVS) is a web-based Master course jointly offered by approximately ten different universities in Europe each year, some of which are campus universities while others are distance education institutions. An extensive description of EVS[1] and its aims is provided in [3].

EVS runs for five months (November 1 till April 1) every year. During that time students work together on sustainability issues in teams of four to seven, with about six to nine teams every year. Ages range between 20 and 25 years for the students from the regular universities and between 30 and 50 years for those from the distance universities. Every team is coached by a tutor and guided by an expert on the team's topic.

The students' final grade for the course can range from 0 to 10 and is comprised of several components: 50% are based on the grade for a team's research report which is given by the expert; 20% are based on the grade for a team's collaboration process which is given by the tutor; 30% are based on the grade for the individual student's contribution which is also given by the tutor. This last grade is called the 'individual-overall' grade (T4) and is divided into three subgrades: 'T1 planning & progress', 'T2 contribution to team' and 'T3 support'. These four grades evaluating an individual student's contribution are the ones used in our analyses. Table 1 explains the different aspects covered by these grades.

Since the run of 2011-2012 EVS has been using an Elgg-based[2] platform which automatically collects and generates data on the students' activities on the platform. This data forms the input to our awareness widget.

[1] http://www.ou.nl/evs
[2] https://elgg.org/

Table 1: Tutor-based grades for students in EVS

	grade	aspects covered by grade
T1	planning & progress	planning a realistic own workload dealing with deadlines and agreements flexibility in making appointments/agreements/planning ability to change roles and responsibilities
T2	contribution to team	dealing with feedback from the group taking initiative, helping the group to progress productivity and quality of contributions
T3	support	being supportive (offering support and help others) encourage the learning of the other members giving feedback / reviewing contributions of others
T4	individual-overall	overall grade (average of the three sub-grades)

Table 2: Calculation of the 5 widget indicator scores

	widget indicator	calculation of the widget indicator scores
W1	initiative	# of posts (discussion, blog, files, pages)
W2	responsiveness	# of comments to posts (discussion, blog, files, pages)
W3	presence	# of page views (on EVS platform)
W4	connectedness	# of contacts made
W5	productivity	(W1 initiative + W2 responsiveness) / W3 presence

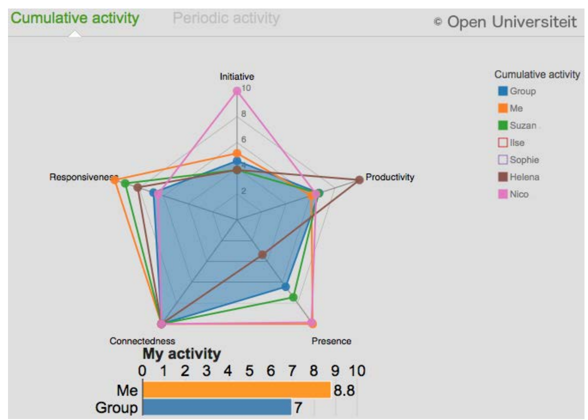

Figure 1: Cumulative student view of the widget.

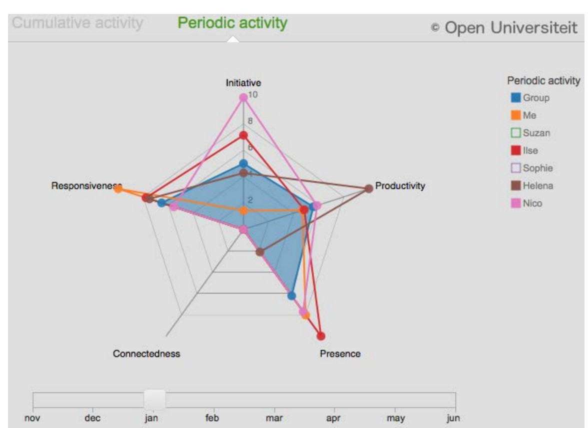

Figure 2: Periodic student view of the widget.

Table 3: Criteria and items of the learner and the teacher section of the Evaluation Framework for Learning Analytics.

		Learners	Teachers
Data	D1:	I know what data is being collected.	I know what data is being collected.
	D2:	I have access to my data.	I have acces to my students' data.
	D3:	I understand the presented results.	I understand the presented results.
Awareness	A1:	I am aware of my current learning status.	I am aware of my learners' current learning status.
	A2:	I comprehend my current learning status.	I comprehend my learners' current learning status.
	A3:	I can project my future learning status.	I can project my learners' future learning status.
Reflection	R1:	I reflect on my learning activities.	I reflect on my teaching activities.
	R2:	I reflect on alternative learning activities.	I reflect on alternative teaching activities.
	R3:	I know when to change my behaviour.	I know when to change my behaviour.
Impact	I1:	I can detect whether I am falling behind.	I can detect whether my students are falling behind.
	I2:	I study more efficiently.	My students learn more efficiently.
	I3:	I study more effectively.	My students learn more effectively.

2.1.2 The Activity Widget

We developed the widget as an Elgg environment plug-in. It can be downloaded under the GNU GPL version 2 [21]. The widget is meant to make students aware of their activities on the platform in relation to those of their team members and to then reflect on this information. It also allows the tutors to become aware of the different activity levels of the students in their team. There are five indicators representing different types of activities on the platform: 'W1 initiative', 'W2 responsiveness', 'W3 presence, 'W4 connectedness' and 'W5 productivity'. Table 2 explains how the scores of the different widget indicators are calculated.

There are two different views available in the activity widget: one showing the widget indicator scores for the whole run of EVS (see Figure 1) and one showing them per month (see Figure 2). The widget indicator scores are automatically calculated from the data recorded in the EVS platform and are scaled from 0 to 10. The team member with the highest activity gets a score of 10 for that widget indicator and the scores of the other team members are then scaled in relation to that. In both views, the team average scores are shown in blue while the current user's scores are shown in orange.

As showing a student's widget indicator scores to the other team members is a privacy sensitive issue, we followed the process suggested by Drachsler and Greller's DELICATE checklist [4] and created a manual explaining the widget's intentions and functionalities. It was distributed to all EVS users making clear what data is collected, how it is visualised and how they can protect their privacy. Implemented within the widget is a Reciprocal Privacy Model (RPM) that allows students to decide whether their team members can see their widget indicator scores or not. Those students that share their data get to see the data from those who also decided to share theirs. Those students that do not want to share their data do not get to see their team members' data. The team average is visible to all students all the time.

2.1.3 The Evaluation Framework for Learning Analytics

The added value of providing learning analytics to students and teachers has clearly been recognised in many educational institutions. While new widgets and dashboards are continuously being developed and implemented, their evaluation has not been standardised yet. We thus developed the Evaluation Framework for Learning Analytics (EFLA)[3] that can be used to evaluate learning analytics tools according to several aspects.

The first version of the EFLA was developed with experts from the learning analytics community using a group concept mapping study [18]. It consisted of five criteria ('Objectives', 'Learning Support', 'Learning Measures and Output', 'Data Aspects' and 'Organisational Aspects') with four items each. In a follow-up study [16], this first version of the EFLA was evaluated by a small group of learning analytics experts. Based on the results of this evaluation combined with a revisit of the original group concept mapping data as well as a thorough look at related literature, a second version of the EFLA was developed. This version is split in two parts, one for learners and one for teachers, that both consist of four criteria ('Data Aspects', 'Awareness', 'Reflection' and 'Impact') with three items each. Table 3 shows the twelve items of the learner as well as the teacher part of the framework. This version was turned into an applicable tool, i.e. a questionnaire for students and teachers, and then used to evaluate the activity widget in EVS.

2.2 Procedure

2.2.1 Correlation and Regression Analyses

As in our exploratory study, we used the scores of the widget indicators 'W1 initiative', 'W2 responsiveness' and 'W3 presence'[4] for our analysis. The other two widget indicators 'W4 connectedness' and 'W5 productivity' were excluded again for the same reasons as in the previous study (see [17]).

We first conducted a t-test to see whether the difference between the widget indicator scores from the online study and those from the exploratory study were significant or not. Then, the scores of the three widget indicators (W1, W2, W3) were correlated with the students' four individual

[3] http://www.laceproject.eu/evaluation-framework-for-la/

[4] For the EVS run of 2011-2012 the 'W3 presence' scores were unfortunately not available.

grades given by the tutors (T1, T2, T3, T4) using Spearman's rank correlation. The ranking corrects for differences in scales and units as well as for differences in grading style of the tutors.

We also applied structural equation modelling in order to determine predictive relations between the widget indicators and the grades. Although the data follows a Poisson distribution because the widget indicators consist of count variables, we could assume a normal distribution because most count variable data had a nearly normal distribution and a mean value far enough from 0[5]. We were thus able to do the regression analysis.

Spearman's rank correlations and the t-test were calculated using IBM's SPSS Statistics 23 while the regression analyses were performed in Mplus 7. All calculations were done for the entire length of the run as well as for the individual months.

2.2.2 *Widget Evaluation*

At the beginning of the course in the fall of 2015, all EVS users received a course manual that included information about the activity widget, i.e. its intentions and functionalities. Two weeks into the course a discussion thread was opened in EVS offering students the opportunity to ask questions about the widget and to comment on it. The discussion thread was kept open and active throughout the course's runtime.

In order to apply the EFLA to the activity widget in EVS, it was turned into a questionnaire. Using online forms, we created a section for each criterion and its three indicators. Every indicator could be rated on a scale from 0 to 6. At the end of the questionnaire, open ended comment boxes were provided for each section asking the users whether they had any comments about this section. Two separate questionnaires were created: one for the students and one for the tutors of EVS.

About halfway through the course, on January 12, 2016, students as well as tutors were sent an invitation to participate in the evaluation of the widget by answering the EFLA questionnaire. They were given ten days to answer. Shortly before the end of the course, on March 18, 2016, students and tutors were invited to participate in a second evaluation round of the widget by answering the EFLA questionnaire again. They were given a week to answer.

3. RESULTS

3.1 Correlation and Regression Analyses

Looking at the average number of actions per student during the different months gives us a first impression of the students' behaviour of the online study in comparison to the data from the exploratory study. Figure 3 shows a student's average number of initiative and responsiveness posts as well as the presence counts per month for the four years of the exploratory study (2011/12, 2012/13, 2013/14, 2014/15) and the year of the online study (2015-2016) where the activity widget was in use. While the number of initiative posts clearly varies a lot between the years, the number of responsiveness posts and presence counts are much closer together.

[5]The mean should be > 10 to be far enough from 0 according to www.umass.edu/wsp/resources/poisson/ and www.umass.edu/wsp/resources/poisson/poisson1.html and www.umass.edu/wsp/resources/poisson/poisson2.html.

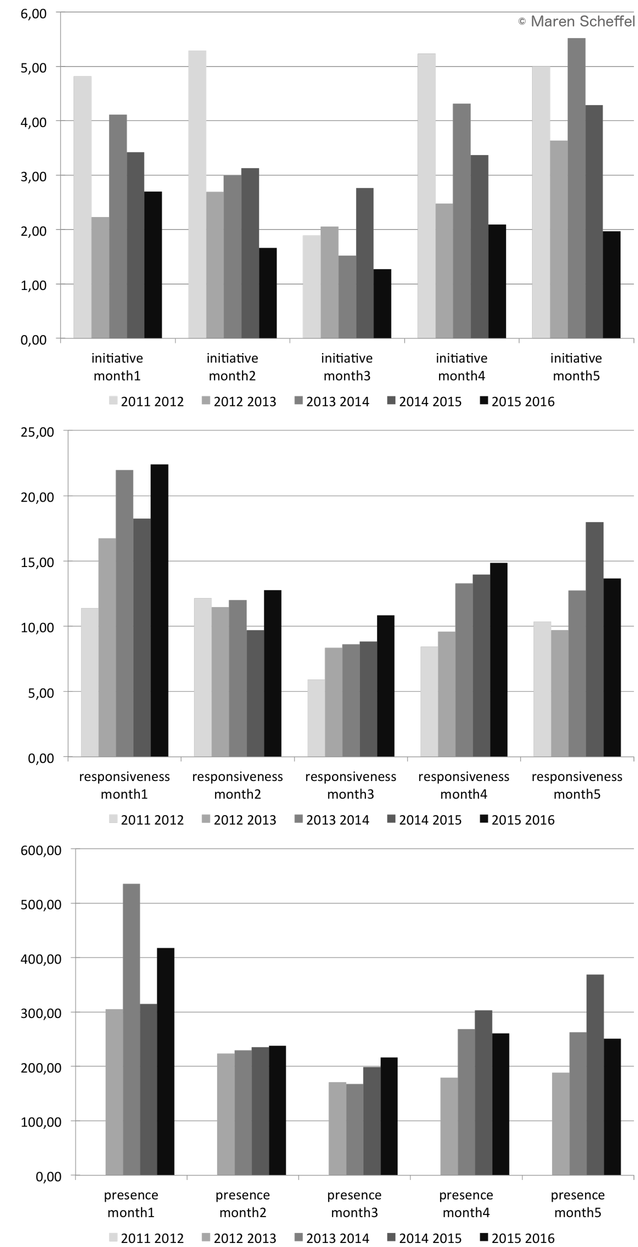

Figure 3: A student's average number of actions for the three widget indicators per month for five different years.

The most striking difference between the years is that the course run with the widget (2015-2016) has the fewest initiative posts (highly significant, $P < 0.000$) and the most responsiveness posts (marginally significant, $P = 0.053$).

The regression analyses were done in two sets: one had T1, T2 and T3 as the dependent variable while the other had T4 as the dependent variable due to T4 being a combination of the other grades. All Root Mean Squared Errors of Approximation and all Standardised Root Mean Squared Residuals were equal to 0.0 while all Tucker-Lewis Indices and all Comparative Fit Indices were equal to 1.0 except for the CFI of the analysis in month3 between the three

Table 4: Spearman correlation coefficients for individual grades (tutor-based) and widget indicator scores (widget-based) based on the individual months from the online study in 2015-2016, n=33.

		W1 initiative					W2 responsiveness					W3 presence				
		m1	m2	m3	m4	m5	m1	m2	m3	m4	m5	m1	m2	m3	m4	m5
T1	Corr.	.336	.154	.188	.200	.297	.421*	.130	.116	.571**	.580**	.221	.024	.119	.453**	.407*
	Sig.	.056	.391	.295	.265	.094	.015	.470	.522	.001	.000	.217	.895	.511	.008	.019
T2	Corr.	.354*	.118	.231	.299	.290	.374*	.101	.274	.599**	.641**	.103	-.120	.018	.393*	.365*
	Sig.	.043	.512	.195	.091	.102	.032	.577	.123	.000	.000	.569	.507	.921	.024	.037
T3	Corr.	.305	.036	.124	.371*	.362*	.331	.039	.149	.641**	.656**	.045	-.142	-.013	.481**	.443**
	Sig.	.084	.844	.491	.034	.039	.060	.830	.407	.000	.000	.805	.431	.942	.005	.010
T4	Corr.	.372*	.098	.174	.306	.342	.378*	.064	.212	.609**	.669**	.146	-.082	.072	.458**	.424*
	Sig.	.033	.586	.333	.083	.051	.030	.723	.237	.000	.000	.416	.650	.689	.007	.014

**. significant at the 0.01 level (2-tailed). *. significant at the 0.05 level (2-tailed).

Table 5: Standardised path coefficients (β) for the individual grades (tutor-based) and the widget indicator scores (widget-based) based on the individual months from the online study in 2015-2016, n=33.

		month 1			month 2			month 3			month 4			month 5		
		W1	W2	W3	W1	W2	W3	W1	W2	W3	W1	W2	W3	W1	W2	W3
T1	β	.228	.499**	-.232	.313	.217	-.335	-.036	-.029	.231	-.314	.388*	.418	-.225	.411*	.284
	Sig.	.208	.004	.208	.128	.253	.110	.909	.879	.441	.177	.028	.074	.389	.046	.346
T2	β	.257	.480**	-.272	.352	.290	-.524**	.110	.150	-.016	-.055	.513**	.125	-.219	.622**	.129
	Sig.	.157	.007	.140	.072	.105	.007	.726	.421	.957	.811	.002	.591	.367	.001	.645
T3	β	.213	.476**	-.267	.160	.206	-.332	-.113	.045	.208	.037	.436**	.185	-.117	.498**	.213
	Sig.	.253	.009	.157	.456	.290	.122	.724	.814	.491	.870	.008	.417	.634	.008	.451
T4	β	.237	.504**	-.261	.288	.246	-.407*	-.017	.054	.154	-.123	.462**	.261	-.194	.526**	.220
	Sig.	.189	.004	.155	.161	.191	.048	.957	.777	.611	.589	.005	.257	.432	.005	.439

**. significant at the 0.01 level (2-tailed). *. significant at the 0.05 level (2-tailed).

Table 6: Spearman correlation coefficients and standardised path coefficients (β) for individual grades (tutor-based) and widget indicator scores (widget-based) based on the entire length of the run from the online study in 2015-2016, n=33.

		correlations coefficients				standardised path coefficients		
		W1	W2	W3		W1	W2	W3
T1	Corr.	.234	.508**	.281	β	.190	.366	-.091
	Sig.	.189	.003	.113	Sig.	.452	.063	.702
T2	Corr.	.285	.518**	.168	β	.299	.500**	-.323
	Sig.	.108	.002	.351	Sig.	.200	.005	.142
T3	Corr.	.266	.512**	.231	β	.214	.404*	-.148
	Sig.	.135	.002	.197	Sig.	.389	.036	.530
T4	Corr.	.285	.527**	.238	β	.238	.438*	-.185
	Sig.	.108	.002	.183	Sig.	.326	.019	.420

**. significant at the 0.01 level (2-tailed).
*. significant at the 0.05 level (2-tailed).

indicators and grade T4 which was equal to 0.0.

In the exploratory study, all grade-indicator combinations except the one between 'T1 planning & progress'/'W3 presence' yielded significant and positive correlations when measuring the students' activity over the entire length of the run. In the online study, however, 'W2 responsiveness' is the only widget indicator that positively and significantly correlates with the four grades (see Table 6[6]). All grade-W2 correlations are significant at the 0.01 level and higher than .500. That is, there are less significant correlations in the online study than in the exploratory study but those that are significant are stronger.

When calculating the correlations for the online study per month instead of the whole run, the results are again quite different from those in the exploratory study. In the exploratory study the scores of the indicators 'W1 initiative' and 'W2 responsiveness' correlated significantly with all four grades in months 1, 2, 3 and 4 with W2 also significantly correlating with the grades T2, T3 and T4 in month5. The indicator 'W3 presence' had the smallest number of significant correlations with the different grades that were rather low. The strongest correlations were obtained between W2 and all grades in month2. Looking at the individual month, the correlation results from the online study with the live activity widget here also look quite different (see Table 4). Overall there are now less significant correlations and hardly any in month1 or month2. The strongest correlation coefficients (ranging from .571 to .669) are received between the 'W2 responsiveness' indicator and the four different grades in month4 and month5. All of them are significant at the

[6]Due to lack of space we only show the online study results (for the correlation as well as the regression analyses). Please refer to [17] for detailed results of the exploratory study.

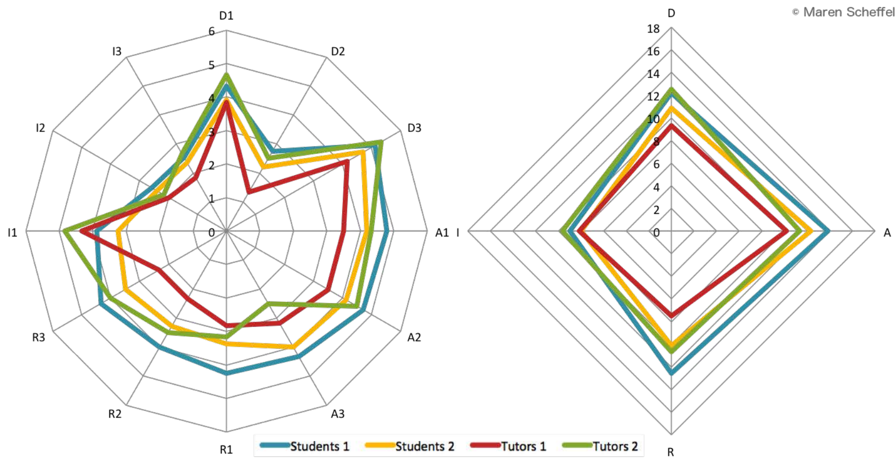

Figure 4: Average scores of the twelve ELFA items on the left and the four criteria on the right for students and tutors for both rounds.

0.01 level. Additionally, the previously low scoring 'W3 presence' indicator now obtains high and significant correlations with all four grades in month4 and month5.

Conducting the structural equation modelling over the entire length of the run in the exploratory study showed that all three widget indicator scores could be used as predictors for all four grades except the 'T1 planning & progress' / 'W3 presence' combination. The 'W2 responsiveness' was the strongest and most significant predictor. In our current online study, there are only three predictive relations (see Table 6), i.e. the 'W2 responsiveness' indicator is a predictor for the grades 'T2 contribution to the team', 'T3 support' and 'T4 individual-overall'. None of the other indicators can be used as predictors.

Comparing the regression analysis results for the individual months from the exploratory study with the online study again reveals a number of differences. Previously the 'W2 responsiveness' indicator was a predictor for all grades in all months with month1 and especially month2 providing the strongest predictive relations. The 'W1 initiative' indicator received a predictive relation with all four grades in month1 and month3 while the 'W3 presence' indicator was negatively predictive for the 'T1 planning & progress' grade only. In the online study, however, the 'W2 responsiveness' indicator can only be used as a predictor in month1, month4 and month5 with the latter one holding the strongest predictive relations (see Table 5). The 'W1 initiative' indicator is in no predictive relation with any of the grades in any of the months. The widget indicator scores of 'W3 presence', though, are in a significant negative predictive relation with the grades 'T2 contribution to the team' and 'T4 individual-overall' in month2.

3.2 Widget Evaluation

In order to gauge how the learners and tutors of EVS evaluate their experience with the activity widget, we asked them to fill out the Evaluation Framework for Learning Analytics (EFLA) questionnaire. As we distributed the questionnaire twice during the course, we are able to compare not only the two user types with one another but also any changes in the users' perception of the activity widget over time. Figure 4 shows the average scores of the twelve questionnaire items as well as the combined criteria for both user types and both rounds.

On average students as well as tutors rated awareness and reflection items higher than the items of the data and impact criteria. Also, while the students on average rated the activity widget more positively in the middle of the course, tutors gave more positive ratings at the end of the course.

Conducting a t-test for the four criteria allowed us to see whether the differences between the two user types or between the two rounds were significant or not. Table 8 shows the mean, standard deviation and standard error mean for the answers given by students and tutors in rounds 1 and 2. We conducted t-tests for four different settings. First, we compared the answers from the students to those from the tutors in round 1 and round 2. We then compared the answers from round 1 to those from round 2 for each user group. Table 7 shows the Levene's test as well as the t-test results for the four different settings.

There are two cases where the differences in ratings are significant. The first one is the rating of the awareness criterion when comparing students and tutors in round 1: $t(28) = 2.158, p = .040$. The second one is the rating of the reflection criterion when comparing round 1 and round 2 of the students: $t(47) = 2.110, p = .040$. None of the other t-tests obtained significant results at the .05 or even the .01 level. In two cases the equality of variance could not be assumed due to the results of the Levene's test. Both of those cases involved the ratings for the reflection criterion from the tutors in round 1, which are rather low, but did not yield significant t-test results. If the equality of variance had been assumed for those cases, however, the difference in ratings between students and tutors for the reflection criterion in phase 1 would have been highly significant (0.006).

From the open ended questions at the end of each EFLA questionnaire we were able to gather some qualitative feedback about the students' and tutors' impression of the activity widget. Generally most students liked the idea behind the dashboard and appreciated to see their platform activities being set in relation to those of their team members. Many students, however, mentioned several issues they were concerned about: The activity widget was not able to reflect activities outside the platform nor did it take the quality of

Table 7: Results of the Levene's tests and the t-tests for four different settings

	Round1: students vs tutors					Round2: students vs tutors					Students: round1 vs round2					Tutors: Round1 vs Round2				
	Levene's test		t - test			Levene's test		t - test			Levene's test		t - test			Levene's test		t - test		
	F	Sig.	t	df	Sig.	F	Sig.	t	df	Sig.	F	Sig.	t	df	Sig.	F	Sig.	t	df	Sig.
D	.209	.651	1.555	28	.131	1.006	.324	-.952	29	.349	.132	.718	1.311	47	.196	.024	.879	-1.078	10	.306
A	2.903	.099	2.158	28	.040	.129	.722	.468	29	.643	2.998	.090	1.350	47	.183	.009	.924	-.376	10	.714
R	4.555	.042	2.236	5.994	.067	2.514	.124	-.273	29	.787	2.889	.096	2.110	47	.040	7.401	.022	-1.327	6.988	.226
I	1.357	.254	.435	28	.667	1.114	.300	-.891	29	.380	1.639	.207	.860	47	.394	1.209	.297	-.572	10	.580

Table 8: Statistics of the EFLA results for students and tutors for both round

		Students				Tutors			
	round	n	Mean	Std.Dev.	St.Er.	n	Mean	St.Dev.	St.Er.
D	1	24	12.21	3.659	.747	6	9.33	5.502	2.246
	2	25	10.84	3.648	.730	6	12.50	4.637	1.893
A	1	24	13.83	3.002	.613	6	10.17	6.014	2.455
	2	25	12.32	4.634	.927	6	11.33	4.633	1.892
R	1	24	12.58	3.296	.673	6	7.50	5.320	2.172
	2	25	10.12	4.720	.944	6	10.67	2.422	.989
I	1	24	9.00	3.901	.796	6	8.17	5.345	2.182
	2	25	8.00	4.223	.845	6	9.67	3.559	1.453

the posts into account. Some students complained that they noticed people posting irrelevant things in order to achieve higher scores. Some students, though, were made aware that they indeed did less than their team mates and thus participated more in the group.

The tutors also expressed their appreciation of the activity widget in the open ended comments and generally liked having the widget as a reference. For most of them, the activity widget confirmed their own impression about their students throughout the course. With regard to the widget fostering reflection about their own tutoring style it was mentioned that such support would be especially useful in those cases where the student groups do not work together well as the tutors could then use the widget to detect such issues early on. One concern the tutors also had was that the activity widget only reflects actions within the EVS platform and that any work the students do with other online tools is not included.

4. DISCUSSION

When student activity is calculated over the whole run of the course, the Spearman correlation results show that in the online study the scores of the 'W2 responsiveness' indicator correlate significantly and positively with the four grades. Research question A1 can thus be answered with a 'yes'. However, while in the exploratory study the scores of all three widget indicators correlated significantly and positively with at least three if not all four of the grades, in the online study only 'W2 responsiveness' did. But although there are now less correlations that are significant, those that are significant are very strong. Regarding the increase of strength of the Spearman correlation results in our online study, similar results are achieved when calculating the activity for the individual months instead of over the whole run of the course. In comparison to the exploratory study, there are also less correlations that are significant in the individual months but those that are significant are very strong.

Going into the online study, we had expected something like this to happen. Of the three indicators analysed, 'W2 responsiveness', i.e. commenting on the posts, pages or files of others, is the one that best represents team interaction and collaboration. With the activity widget in use during the course, we expected it to foster the students' awareness and reflection about their position within the team and the team as a whole and to thus facilitate collaboration processes.

The standardised path coefficients from the structural equation modelling (see Tables 5 and 6) show that there are indeed significant predictive relations between some of the widget indicator scores and the grades. As with the correlation coefficients, only the 'W2 responsiveness' scores receive significant results when looking at the entire length of the run. Slightly more diverse results become apparent when looking at the standardised path coefficients for the different months, e.g. in month2 the score of the 'W3 presence' indicator can also be seen as a predictor for some of the grades. However, the best and by far the most frequent predictor for all four grades are the scores from the 'W2 responsiveness' indicator. Research question A2 can thus also be answered with a 'yes'. Since the 'W2 responsiveness' indicator scores, although surprisingly at the time, had been by far the best predictor in the exploratory study and since we had anticipated an increase of team interactions due to the widget triggering awareness and reflection processes, we had expected this indicator to be the best overall predictor in the online study as well.

What surprised us, however, was that in the online study none of the predictive relations involved the 'W1 initiative' indicator. As presented earlier, the 2015-2016 run had a significantly lower number of initiative posts per student. When looking into the log data from this year it became apparent that the number of posted files (which is by far the major contributor to the initiative score) was lower during the year of the online study. This may be explained by an increased use of external tools already from early on in the course which cannot be logged and was thus not included in the calculation of the widget indicator scores.

During the exploratory study this use of external tools, especially during the second half of the course where the different group reports needed to be written, turned out to be the most likely explanation for the widget indicators of the first half of the course to be better predictors than those of the second half. As the students in the online study also made use of such external tools, we expected the same to be true for the 2015-2016 run even though the widget was now in use. However, research question A3 has to be answered with a 'no' as the strongest correlations and best predictive relations are now more likely to happen towards the end of the course.

This shift to the last few months now being the main source for widget scores with predictive power is already indicated by the correlation results: all grade / W2 as well as all grade / W3 combinations in month4 and month5 are significantly and positively correlated with correlation coefficients ranging from .365 to .669. Month2 and month3 do not show any significant correlations in the online study whereas they did so for many grade / widget indicator combinations in the exploratory study. With regard to predictive relations, while none of the widget indicator scores from month3 can be used as a predictor for any of the grades, there are two predictors in month2: The scores of the 'W3 presence' indicator are in a significant negative predictive relation with the grades 'T2 contribution to team' and 'T4 individual-overall'. With regard to the best predictor, the results of the regression analysis in the online study confirm the afore mentioned shift and show that for three grades the best predictors are the scores of the 'W2 responsiveness' indicator in month5, except the grade 'T1 planning & progress' that is best predicted by the 'W2 responsiveness' indicator score in month1.

In the previous years, the widget indicator scores in the last months were poor predictors of the grades, which we attributed to the students mostly using non-logged external tools in this period. In the online study, the widget indicator scores in the last months were the best predictors of the grades. We can think of two probable causes of this shift. First, students that were initially less active may have been stimulated by widget feedback to become more active, resulting in better grades. Second, students, aware that their activities with the external tools were not captured by the activity widget, posted more frequently on the EVS platform as they wanted the widget to reflect their being active in the course.

Another surprising observation for us was the students' neglect of the privacy protection option through the reciprocal privacy model. None of the students disabled this functionality to mask their data from their team. This could be due to the nature of the collaborative learning process that requires to be aware of the status of other students. In fact, we received generally positive responses from the students about the activity widget and that it indeed supported their team awareness processes as well as added a 'fun factor' to the online learning environment.

The results of the formal evaluation of the activity widget using the EFLA questionnaire show that the answer to research question B1 is 'yes, but for the students' reflection criterion only' as their reflection ratings in round 2 were significantly lower than those in round 1. In all other cases, neither for the students nor the tutors was there a significant difference in evaluation results between the two rounds. From what we were able to gather from the open ended questions as well as the discussion thread, this difference was most likely due to the students feeling less accurately represented the more the course progressed as the activities of the external tools was not reflected in the widget scores. When comparing the evaluation results from the two user groups with one another, the only significant difference is that of the awareness criterion in round 1. Here, students have rated the awareness items significantly higher than the tutors did. In all other cases, neither in round 1 nor in round 2 was there a significant difference in evaluation results between the two user groups. Research question B2 can therefore be answered with 'yes, but for the awareness criterion of phase 1 only'. This is most likely due to the generally positive reception of the activity widget by students already at the beginning of the course while tutors used and thus appreciated the widget more towards the end of the course when they saw their personal impressions about the students confirmed. Except for those two cases, students and tutors thus evaluated the activity widget in a very similar way.

Combining the EFLA results with the comments gathered via the open ended questions allows us to conclude that both students and teachers generally liked and appreciated the activity widget and felt supported in their awareness and reflection processes. Both user groups, however, had issues with the widget's data access (D2) as well as its support of more efficient (I2) and more effective (I3) learning. Additionally, both user groups found it problematic that the activity from external tools could not be included in the widget. Students would also like to see not only the quantity but also the quality of their discussion posts to be taken into account as they otherwise fear that too many irrelevant message are posted to increase the widget indicator score.

We had already identified the risk of students 'playing the system' during our exploratory study and had thus provided a detailed user manual at the beginning of the 2015-2016 course explaining the activity widget's aim and functionalities. This, although being an important step, however, does not seem to have been enough. As emphasised in other studies [12, 1, 10] learning analytics visualisations need to be tightly embedded into a course's instructional design, especially if they are to be used by the students themselves. For the next run of EVS we will therefore carefully take the gathered results into account in order to improve the activity widget as well as the instructional design and to enhance the user experience.

There are several limitations of our study. Due to the change in student population, the students' behaviour in the five different runs cannot be set into a one-to-one relation. Their previous experience with and usage of online learning platforms as well as external communication and collaboration tools influences the cohort's actions. The same applies to the tutors. Although many of them have been tutors for EVS for a number of years, their experience and interactions with their student groups also changes from year to year. Related to this aspect of change in student population, student and tutor behaviour as well as external tools is another aspect that has to be kept in mind when looking at the results of our online study: although a number of our observations can be explained as effects of the activity widget being in use, there is no proof that this is the case. Only after observing and analysing further years of the EVS will we be able to attribute differences between the years that did not have the widget and those that did clearly to the use of the widget.

5. CONCLUSIONS

This paper presented an empirical study conducted with data collected during the five months of a live Master course where students work collaboratively in virtual teams. We implemented a learning analytics-based activity widget to foster awareness and reflection among the team members into the course's online learning platform and examined the predictive power of the widget indicators towards the students' grades of this course in comparison to the data from previous years where the widget had not been in use. Our results indicate that the widget indicator 'responsiveness', i.e.

the number of response posts made on the course's platform, is a significant positive predictor towards the grades. In the years without the widget, the students' behaviour of the first few months of the course held more predictive power, whereas in the year where the widget was implemented into the platform, the last few months of the course had a higher predictive potential. This, in combination with the results from a quantitative as well as qualitative evaluation of the activity widget during the course, suggests that the differences between the years could be explained by the use of the widget and its effective fostering of awareness and reflection. More investigations are needed in order to provide further evidence that can substantiate this hypothesis and confirm the effectiveness of the widget. We will therefore continue to deploy the activity widget in future editions of the course.

6. ACKNOWLEDGMENTS

This work was partly funded by the LACE project (GA No. 619424) under the FP7 programme of the European Commission.

7. REFERENCES

[1] S. Beheshitha, M. Hatala, D. Gasevic, and S. Joksimovic. The role of achievement goal orientations when studying effect of learning analytics visualizations. In *Proc. of the 6th Int. Conf. on Learning Analytics & Knowledge*, LAK '16, pages 54–63, New York, NY, USA, 2016. ACM.

[2] D. Butler and P. Winne. Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research*, 65(3):245–281, 1995.

[3] J. de Kraker and R. Cörvers. European Virtual Seminar on Sustainable Development: international, multi-disciplinary learning in an online social network. *E-learning and Education for Sustainability, Series 'Environmental Education, Communication and Sustainability'*, 35:117–136, 2014.

[4] H. Drachsler and W. Greller. Privacy and Analytics - it's a DELICATE issue. A Checklist to establish trusted Learning Analytics. In *Proc. of the 6th Int. Conf. on Learning Analytics & Knowledge*, LAK '16, pages 89–98, New York, NY, USA, 2016. ACM.

[5] M. R. Endsley. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, 37:32–64, 1995.

[6] M. R. Endsley. Theoretical underpinnings of situation awareness: a critical review. In M. R. Endsley and D. J. Garland, editors, *Situation Awareness Analysis and Measurement*, pages 3–28. Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2000.

[7] R. Ferguson and D. Clow. Learning Analytics Community Exchange: Evidence Hub. In *Proc. of the 6th Int. Conf. on Learning Analytics & Knowledge*, LAK '16, pages 520–521, New York, NY, USA, 2016. ACM.

[8] M. Fishbein and I. Ajzen. *Predicting and Changing Behavior: The reasoned action approach.* Psychology Press, New York, NY, USA, 2010.

[9] J. Hattie and H. Timperley. The Power of Feedback. *Review of Educational Research*, 77(1):81–112, 2007.

[10] I. Khan and A. Pardo. Data2U: scalable real time student feedback in acive learning environments. In *Proc. of the 6th Int. Conf. on Learning Analytics & Knowledge*, LAK '16, pages 249–253, New York, NY, USA, 2016. ACM.

[11] P. Kirschner, K. Kreijns, C. Phielix, and J. Fransen. Awareness of cognitive and social behaviour in a CSCL environment. *Journal of Computer Assisted Learning*, 31(1):59–77, 2015.

[12] S. Lonn, S. Aguilar, and S. Teasley. Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in Human Behavior*, 47:90–97, 2015.

[13] L. McAlpine and C. Weston. Reflection: Issues related to improving professors' teaching and students' learning. *Instructional Science*, 28(5):363–385, 2000.

[14] E. H. Mory. Feedback Research Revisited. In D. H. Jonassen, editor, *Handbook of Research on Educational Communications and Technology*, pages 745–783. Lawrence Erlbaum Associates, Mahwah, NJ, US, 2004.

[15] C. Phielix, F. Prins, P. Kirschner, G. Erkens, and J. Jaspers. Groups awareness of social and cognitive performance in a CSCL environment: Effects of a peer feedback reflection tool. *Computers in Human Behavior*, 27:1087–1102, 2011.

[16] M. Scheffel, H. Drachlser, and M. Specht. Developing an evaluation framework of quality indicators for learning analytics. In *Proc. of the 5th Int. Conf. on Learning Analytics & Knowledge*, LAK '15, pages 16–20, New York, NY, USA, 2015. ACM.

[17] M. Scheffel, H. Drachsler, J. de Kraker, K. Kreijns, A. Slootmaker, and M. Specht. Widget, widget on the wall, am I performing well at all? *IEEE Transactions on Learning Technologies*, PP(99):1–1, 2016.

[18] M. Scheffel, H. Drachsler, S. Stoyanov, and M. Specht. Quality Indicators for Learning Analytics. *Educational Technology & Society*, 17(4):117–132, 2014.

[19] D. Schön. *The reflective practitioner: How professionals think in action.* Temple Smith, London, UK, 1983.

[20] G. Siemens, S. Dawson, and G. Lynch. Improving the Quality and Productivity of the Higher Education Sector - Policy and Strategy for Systems-Level Deployment of Learning Analytics. SoLAR report for the Australian Government Office for Learning and Teaching, December 2013.

[21] A. Slootmaker, M. Scheffel, K. Kreijns, J. De Kraker, and H. Drachsler. Performance dashboard to support awareness and reflection in elgg communities (version 1.15) [software], 2015.

[22] K. Verbert, E. Duval, J. Klerkx, S. Govaerts, and J. Santos. Learning Analytics Dashboard Applications. *American Behavioral Scientist*, 57(10):1500–1509, 2013.

[23] P. H. Winne. Inherent Details in Self-Regulated Learning. *Educational Psychologist*, 30(4):173–187, 1995.

[24] P. H. Winne. How Software Technologies Can Improve Research on Learning and Bolster School Reform. *Educational Psychologist*, 41(1):5–17, 2006.

[25] B. J. Zimmerman. Self-Regulation Involves More Than Metacognition: A Social Cognitive Perspective. *Educational Psychologist*, 30(4):217–221, 1995.