Measuring and visualizing students’ behavioral engagement in writing activities

Ming Liu, Rafael A. Calvo, Abelardo Pardo, and Andrew Martin

Abstract—Engagement is critical to the success of learning activities such as writing, and can be promoted with appropriate feedback. Current engagement measures rely mostly on data collected by observers or self-reported by the participants. In this paper, we describe a learning analytic system called Tracer, which derives behavioral engagement measures and creates visualizations of behavioral patterns of students writing on a cloud-based application. The tool records the intermediate stages of document development and uses this data to measure learners’ behavioral engagement and derive three visualizations. Writers (N = 23 University students) participated in a controlled one-hour writing session in which they post-facto self-reported their level of behavioral engagement. Results show that the level of behavioral engagement automatically estimated by the system correlates with the level reported by the participants. Additionally, users stated that the visualizations were coherent with their writing activity and were useful to help them reflect on the writing process.

Index Terms—E-Learning Tools, Computers and Education, Visualization

1 INTRODUCTION

The use of technology in educational environments has grown significantly in the past few years, and with it a growing awareness of the importance of student engagement [1] as well as the capacity for technologies to capture this engagement information. Traditionally, learning technologies have been used to support administrative tasks (e.g. access to course notes and assignment submission), however more recently, the systems’ capacity to collect large amounts of data about students’ behavior is being harnessed to improve learning interactions and to personalize the learning experience [2].

There is a large body of literature examining the behavioral and psychological factors that affect learning. Using computers, there is now the capacity to ‘observe’ students ‘in situ’, that is, while students are occupied in learning activities [3]. Rather than focusing on outcome of an assessment alone, computers can now observe students while they participate in an activity. This capacity has provided an ideal foundation to explore learning processes and their cognitive, affective and behavioral components. There is good evidence [4] to suggest that a student who is engaged and intrinsically motivated in a task is more likely to learn from an activity and models of school engagement identify three core dimensions: behavioral, cognitive and emotional engagement.

‘Behavioral engagement’, which is the focus of the present study, refers to participation in school related activities and involvement in academic and learning tasks such as those being done online. It can be measured by observation and self-report. ‘Cognitive engagement’ refers to motivation, thoughtfulness and willingness to make an effort to comprehend ideas and master new skills. ‘Emotional engagement’ includes emotions and interest, such as affective reactions in the classroom towards teachers. These three aspects are interrelated and helpful to understand engagement as a whole. The ‘engagement’ used throughout the paper refers to ‘behavioral engagement’ unless we specify it as ‘cognitive engagement’ or ‘emotional engagement’.

Compared with emotional and cognitive engagement, the measurement of behavioral engagement is more straightforward because behavioral patterns can be defined, observed and interpreted. For instance, when a student participates in an activity that is technology mediated, a detailed collection of behavioural events can be recorded. Computer keystroke logging [5, 6] or screen capturing [7] allow a detailed account of the behavior of a writer including actions such as starting a new paragraph or deleting a text portion and these are all considered indicators of behavioural engagement. Thus, new computer technology permits the observation and identification of learning events, which can then be examined in relation to other indices of engagement. However, these technologies require specialised setups and often hardware. These factors present a barrier to the use of this technology in the education sector.

New cloud-based technologies, such as Google Docs not only record the revision history (each revision contains the content and timestamp) they also provide programming API to access this information. In addition, Google Docs has the advantage of supporting easy system integration and synchronous collaborative writing and it has been successfully applied in student assignment management [8] and collaborative writing practices [9].
We report the development and evaluation of a new method to measure and visualize student behavioral engagement that was trialed with University students. In this study participants were required to complete writing tasks while their reading and writing activities were recorded using facilities that extend cloud-based writing tools. Computer-generated observations were processed and visualizations generated to yield estimations of the writer’s level of engagement. The computer-generated estimations of engagement were compared with self-reported levels of engagement in order to evaluate the concordance between these two measures as well as to determine if the point-based and line-based visualizations were useful to reflect the overall writing process.

The major contributions described in this paper are: 1) a novel learning analytic system that collects behavioral data of users writing, estimates the level of engagement, and generates three types of visualizations, point-based, line-based and height-based visualizations; the study also examined 2) the concordance of the inferred engagement measures by comparing these with participant self-reports.

The remainder of this paper is organized as follows: Section 2 describes the relevant work relating to behavioral engagement and learning analytics. Section 3 describes the architecture of the system used in the study. In section 4, the algorithms used to process the engagement measurements and the creation of the three types of visualizations are described. Sections 5 describe the research scenario and experimental study used to validate the proposed approach. The paper concludes in Section 6 with a discussion of the overall approach as well as lines for future exploration.

2 BACKGROUND

2.1 Behavioral Engagement

Studies of behavioral engagement in learning environments typically use evidence collected by human observers, such as teachers or students [10, 11]. For example, using scales such as the Student Engagement Walkthrough Checklist, observers such as administrators, instructional supervisors or teachers, have examined the degree to which students exhibit engagement in the classroom, by measuring behaviours such as positive body language, consistent focus, verbal participation, as well as confidence, enjoyment and excitement [12]. The observer ratings are then compared to simultaneous and anonymous ratings by students of their level of engagement according to the extent to which the work is interesting and challenging, and the degree to which they understand why and what they are learning.

Jones [12] have defined the models of general engagement including behavioral, emotional and cognitive engagement as consisting of three dimensions; intensity, consistency and breadth. Intensity relates to the level of engagement of each student. Consistency refers to how long students remain engaged at high levels throughout the class period and breadth refers to how broadly the class as a whole is engaged. Measuring dimensions of engagement allows teachers to provide differentiated feedback. For example, if the engagement intensity is low, teachers can focus on adding rigor and relevance to expectations and lessons.

To date, most of the research on student engagement has occurred in classrooms [13], yet researchers are increasingly exploring learning theories in web-based activities [1], social software [14], smart interactive devices [15] and virtual environments [16]. ‘Clickers’ [15] allowed students to quickly answer questions presented in class. Responses can be anonymised or identified and software programs are usually used to summarize responses and present visualisations in the form of charts. Technology-based tools such as Wiki technology [14] have been used to support learning engagement. Cole [14] tested Wikis in a third year undergraduate course to examine the degree to which they supported student knowledge construction, peer interaction and group work. However given the optional nature of this form of technology in the course, students did not contribute to the Wiki as was intended. Thus focus groups were used to examine barriers to uptake rather than the affects of Wikis on student engagement per se. However, a limitation of previous studies is that they have not addressed how to automatically track and analyze student behavior patterns and present them in a way that is understandable. Given the difficulties identified by previous studies [14] related to student use of web-based techniques the present study was trialed within a laboratory environment rather than as part of a course.

The present study sought to implement Jones’ [12] intensity and consistency measures of engagement using an automated analytics system and evaluate its accuracy to automatically measure engagement in the context of writing activities.

2.2 Learning Analytics

The area known as Learning Analytics (LA) has emerged as a result of behavior-related information available about how students learn. LA is defined as “the measurement, collection, analysis and reporting data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs”[17]. In general, learning analytic systems can be divided into several modules, steps or phases [18]. One module captures detailed events such as the number and frequency of interactions with resources in a learning management system[19]. This module may also use additional factors such as a student’s Grade Point Average (GPA) [20], gender, etc. An algorithmic module then analyzes these data to infer some conclusions. These conclusions are reported back to users through an additional module. Typical reports include visualizations that can range from a simple traffic light-like display of overall student status and risks [2, 21], to more sophisticated dashboards with detailed information about various aspects derived from the data [22-24]. A more advanced module is often incorporated to suggest actions to modify the learning behaviors. These actions are sometimes referred to as interventions and may range from suggestions automatically proposed to in-
The initial analytics tools designed within the context of learning experiences were pedagogically neutral. In other words, they simply provided insight about the events occurring in an environment without targeting any specific strategy. An example of these early tools is CourseVis[25], a platform to visually represent the interactions of students in the context of web-based distance education. The concept of dashboard appeared as a proposal to centralize the visualization of student events and foster self-reflection and sensemaking for both students and teachers[26]. At the same time, academic institutions started to use analytics to tackle the problem of student retention. Numerous institutions have created platforms that combine student interactions with other socio-economic factors to calculate the probability of a student dropping a course[2]. These platforms were later extended to cover the anticipation of other facts such as academic performance and are generally known as Early Warning Systems or EWAs (see [27, 28] for two examples of these systems).

In the recent years, the influence of pedagogical intent has been gaining influence in the design of learning analytics approaches. The emergence of constructivist approaches to education prompted the appearance of applications to analyze the interaction of users within the context of social networks. The work of Aviv et al. showed how to connect the topology emerging in a network with knowledge construction[29]. Visualizations are also used in this context to identify specific patterns and promote a more cohesive network [30]. More advanced approaches have been recently proposed in the context of discourse analysis[31-33]. Nowadays, an increasing number of techniques and contexts are included as learning analytics techniques such as process mining, recommendation algorithms, text analytics, resource analysis, etc.

The system described in this document can be categorized as a visualization and self-reflection tool. The system measures the level of engagement of users while participating in a writing task and then creates visualizations of this engagement to promote reflection.

3 System Architecture

Our learning analytics system ('Tracer') captures a detailed account of how learners engage in a writing activity, estimates the user’s engagement, and produces three types of visualizations (graphs) of this engagement. The three components of the system are shown in Figure 1. The first is the Data Collection Module which currently relies on the combination of a Google Doc API and iWrite [8]. iWrite is used to handle assignments, manage how they are shown to users, and save the final version of the writing task as a PDF document, generally stored to be assessed by the instructor. Google Docs is used as the supporting editor. The application records numerous intermediate versions of the documents while they are being modified. The application programming interface (API) is used by Tracer to access this sequence of documents. Initial validation of Tracer was completed using data generated from University students participating in a writing activity [34]. Tracer also obtains from iWrite additional writing activity parameters such as start time, end time etc.

The second component of Tracer is the Data Analysis Module in which two engagement measurement algorithms are implemented described in section 4.1. The third component is the Feedback Module described in section 4.2 where visualizations are created based on the results derived from the analysis phase [34].

4 Engagement Measurement Algorithms and Visualizations

Due to the complexity of the data captured during the writing activity, it is challenging to produce a simple and meaningful visualization. Thus, raw events data are analyzed by using different engagement measurement algorithms and visualizations. This section describes two engagement measurement algorithms, point-base and intensity-based algorithms, and three visualizations: Point-based Visualization (PbV), Line-based Visualization (LvV) and Height-based Visualization (HbV). The objective of this component was to explore how best to convey
information to the user in an understandable format [34].

4.1 Engagement Measurement Algorithms

The point-based algorithm simply sum up each data cluster, where each data cluster has a fixed threshold, such as 1 minute, while the intensity-based algorithm sum up several weighted series of data points, where each series has a weight referred to the intensity of adjacent data points within the series.

4.1.1 Point-based Engagement Measurement Algorithm (PbA)

calculateEngagementForPoint(List events, float threshold, float scale)

- Date startTime = Time of first event
- Date currentTime, startSegment;
- Int clusterCount = 1;
- Float duration, segmentDuration;
- startSegment = startTime;
- For each event in the list
  - currentTime = time of the event;
  - segmentDuration = Duration(startSegment, currentTime);
  - duration = Duration(startTime, currentTime);
  - if (duration > threshold || segmentDuration > threshold)
    - clusterCount++;
    - startSegment = currentTime;
  - startTime = currentTime;
  - return clusterCount * scale;

Figure 2. Point-based Engagement Measurement Algorithm

The PbA is based on the number of data point clusters. A data point cluster is a set of consecutive events separated in time by less than a certain threshold. This threshold is calculated depending on the intensity required for the writing task. For example, if the writing task is intensive, we need a smaller threshold to detect the clusters. By default, the threshold is 1 minute while the scale is minute. The final engagement score is calculated with the following equation and algorithm.

Engagement=ClusterCount*Scale

Where the ClusterCount is the number of clusters and the Scale could be a minute or an hour. The pseudocode for the algorithm to compute the engagement is illustrated in Figure 2.

4.1.2 Intensity-based Engagement Measurement Algorithm (IbA)

While the PbA simply calculate the number of data cluster the intensity-based algorithm (IbA) incorporates the intensity of engagement and assigns a different weight to each data cluster/series according to their intensity. The IbA is thus proposed as an alternative method to compute engagement. In this algorithm, a series is defined as a group of events represented by a line. These events are grouped based on the duration between neighboring events. Each line is associated with a weight that indicates the intensity of the line. Hence, the whole graph is made of lines. The weighting process is defined as follows:

1. We define a hashmap, where each entry contains a time threshold and a corresponding weight value. For example, (0.5h, 0.8) indicates that the time threshold is 0.5h and its corresponding weight is 0.8.

2. If the duration between neighboring events is less than the smallest time threshold, we assign that corresponding weight to the series to which that segment belongs. Based on our experiences with writing activities in learning situations we considered the following combinations/hashtags: (0.5h, 1), (1h, 0.8), (3h, 0.4) and (12h, 0.2). For example, if the duration of an activity is 2 hours, we assigned 0.4 as a weight to the series because 3h is the smallest value defined in the hashmap that is larger than 2h. In a one month project proposal writing assignment.

Thus the total engagement score is calculated as the following weighted sum:

$$\text{Engagement} = \sum_i s_i \times w_i$$ (2)

calculateEngagementForLines(List events)

- Date startTime = Time of first event;
- Int startSeriesID = The seriesID belongs to this event;
- Date currentTime;
- Int currentSeriesID;
- Float duration, score;
- Int seriesID = The seriesID which this event belongs to;
- For each event in the list
  - currentTime = time of the event;
  - currentSeriesID = the seriesID belongs to this event;
  - currentSeriesID = the seriesID belongs to this event;
  - if (seriesID = currentSeriesID)
    - duration = Duration(startTime, currentTime);
    - score += duration * getHashWeight(durationOfSeries);
    - startSeriesID = currentSeriesID;
    - startTime = currentTime;
  - if (lastInSeries(startEvent),
      - calculate the engagement and add it to the total score
    - if (lastInList(startEvent),
      - calculate the engagement and add it to the total score
  - return score;

Figure 3. Intensity-based Engagement Measurement Algorithm
where \( i \) is the index of a series, \( S_i \) is the duration of the series \( i \) and \( W_i \) is the weight assigned to \( i \). The pseudocode of the algorithm to compute the engagement is shown in Figure 3.

### 4.2 Visualizations

The two algorithms above explore different ways to measure student engagement and engagement intensity based on the information derived from the raw data events. The purpose of these visualizations is to help students easily check their engagement in an activity, reflect on their behaviors, and change it if not fully engaged. This section presents different visualizations produced with these measures. The configuration of the visualization is dependent on the set up of a writing activity. For instance, a writing activity can be divided into multiple sub-activities, such as initial drafting, reading feedback and revising.

#### 4.2.1 Point-based Visualization

In the Point-based Visualization (PbV), each row represents a user’s behaviour and each point represents the user action at a particular time, such as drafting an initial version or revising a document. The visualization can represent multiple participants at a time. Figure 4 shows the behavioral pattern of an engineering student working on a writing activity (project proposal assignment) within a month using a PbV. In this case, the activity is divided into 4 sequential sub-activities described in the activity’s timeline: initial drafting, peer reviewing, reading feedback given by peers, and writing the final version. The vertical lines indicate submission deadlines for each sub-activity. The activity information was obtained from iWrite including the deadline of each sub-activities\(^\text{[34]}\). Tracer reads raw data from iWrite, and then generates graphs. Thus, PbV generation does not require any engagement measurement algorithm since PbV is just a

![Figure 4. Point-based Visualization: a data point represents a user action during the writing process. For example, a green circle represents a user action during the drafting process, a blue diamond during the peer-reviewing process, a red square during a feedback reading process while a green plus during the revising process. This graph is copied from [34].](image)

---

#### 4.2.2 Line-based Visualization

Figure 5a. Line-based Visualization: a line represents a series of continuous user actions during the writing process. Lines with different colors represent a user’s actions in different writing processes.

![Figure 5a. Line-based Visualization: a line represents a series of continuous user actions during the writing process. Lines with different colors represent a user’s actions in different writing processes.](image)

---

Figure 5b. Line-based Visualization: green lines with different thickness show that a user has done several intensive writing in the drafting process. Graphs are copied from [34].

![Figure 5b. Line-based Visualization: green lines with different thickness show that a user has done several intensive writing in the drafting process. Graphs are copied from [34].](image)
timeline of events.

### 4.2.2 Line-based Visualization

The Line-based Visualization (LbV) uses a line to connect the points and the thickness of a line indicates the intensity of the user's behavior during a period of time. This information is derived from IbA, where a series represents a line and its weight represents a line thickness (see Figure 5).

Figure 5a shows the behavioral pattern of an engineering student whilst completing a writing activity where the points are connected into lines in each sub-activity. We defined each line as a series, which has a different color and thickness. Figure 5b is a supplementary figure (zoom in of Figure 5a), which gives more detail about the student’s behavior during the writing task. As can be seen, this visualization includes not only how the events occur during the experience (as was the case in PbV) but also the level of intensity.

### 4.2.3 Height-based Visualization

In the Height-based Visualization (HbV), a rectangle represents a series of user actions; where its width represents the timeline and its height shows the intensity of engagement. This information is derived from IbA, where the width is a series and height is its weight. Like the LbV, HbV can show engagement consistency and intensity. Figure 6 shows the engagement of a student in a writing activity that contains only drafting and revising sub-activities within a one-hour time frame. The green bar shows the drafting process, and the blue bar shows the revising process. In this type of visualization, the engagement intensity and consistency are illustrated; for instance the period of high engagement during the drafting phase from 10:14 to 10:42 is clearly depicted. Figure 6 is generated from a one-hour writing activity where Figure 4 and Figure 5 are generated from a one-month writing activity.

In short, three visualizations show user actions during the writing process. We use different colors to represent different writing process. In the line-based and height-based visualization, a series of continuous user actions with an engagement level is represented. The engagement level for each series is derived from the IbA. In the line-based visualization, each line represents a series of user actions and its thickness represents the engagement level. In the height-based visualization, each rectangle represents a series of user actions and its height represents the engagement level. In the Point-based visualization, one single user action is represented as a point.

### 4.3 High frequency revision sampling

The second component to this study extends our previous research in learning analytics and student engagement. We previously conducted a user study where 38 student participants who were enrolled in a fourth year engineering course were required to write a project proposal over one month. Tracer used the information obtained from Google Doc and iWrite to track different writing activities (e.g., drafting proposal, peer reviewing, finalizing proposal). Based on the information obtained, we gauged engagement time and generated two types of visualizations: PbV and LbV, for each student to then evaluate. The results indicated that students generally understood the visualization, were somewhat neutral as to the usefulness of the visualizations for self-reflection, but generally agreed with what the visualizations were showing about their engagement.

Google Doc shows a convenient way to track students’ writing behaviors. However, two main issues were found in our previous study. First, Google Doc API presents difficulties in obtaining accurate recordings of writing behavior. For example, only a small portion of the document revision history could be retrieved. Second, the cloud-based approach cannot track users’ writing behaviors if they ‘copy and paste’ from another editor. These two issues adversely influenced the quality of visualiza-

---

**Figure 6.** Height-based Visualization: a series of continuous user actions is represented as a rectangle where its height means the engagement intensitivity while its width means the consistency. A green rectangle means a user’s continuous action in the drafting process while a blue rectangle means a user’s continuous action in the revising process.
The present study sought to redress these issues. In this study, we improved our methods for recording revision logs: a Google document was embedded in a web page of Tracer and a Javascript function was implemented in that page to regularly (every five seconds) call Google API to download the latest document revisions while the web page was open. In this way, we sought to overcome the limitation of the Google Doc API. In addition, we evaluated a new visualization method: height-based visualization.

5.1 Participants and Procedure
A total of 23 university students participated in this study. The participants’ age ranged from 20 to 60 years (M: 34, SD: 11) and there were 14 males and 9 females. Those student participants came from different disciplines, including engineering (14) and education (3). They had no prior knowledge of Tracer and did not participate in any previous related study. They signed an informed consent form approved by the University's Human Research Ethics Committee.

We arranged a separate one hour writing activity for each participant. The writing activity shown in Figure 7 included a drafting session (30 min) and a revising session. After completing the draft, participants were asked to wait for their feedback. Within 10 minutes they received feedback and started revising the document. We conducted this study in a controlled environment so that each participant could only write using an assigned Google document embedded in Tracer (see Figure 8), thus avoiding the ‘copy-and-paste’ issues in the previous study. Once the participants finished the draft, generic feedback was provided to them. In this study, an experimenter emailed the same predefined feedback to each participant to read after the drafting stage. An example of feedback is shown below:

![Figure 7 Writing Activity Work Flow](image)

![Figure 8. A screenshot of writing a travel experience in Tracer](image)
- Highlight the best experiences from your perspective. Long trips can present a lot of information to cover and writing about day to day activities is not always important to fellow travellers. Keep the article focused on the important points of the trip.
- Add pictures to your review that allow you to show the experience as you tell about it in your writing. It is possible to paint a picture with words, but readers will most likely want to see pictures of your experiences as well.

After the writing activity was finished, Tracer generated 3 visualizations (PbV, LbV, HbV) and 2 engagement measurements for each participant derived from the Google document revisions. At the conclusion of the experiment, each participant was asked to estimate their engagement time in the writing session and rate the three types of visualizations using a Likert scale, where 1 was “strongly disagree” and 5 was “strongly agree”. Participants were asked the following quality measure (QM) questions:

QM1: I understand what the visualization is trying to convey.
QM2: I agree with what the visualization is showing.
QM3: The visualization is useful for me to reflect on what I did.

5.2 Results
Table 1 illustrates the average scores reported by participants to the three visualization types. The average quality measure scores QM1 and QM2 were above 4, indicating that most participants agreed that they understood what the visualizations were trying to convey. PbV and HbV obtain higher scores than LbV, in QM1 (PbV: 4.26, LbV: 4.04 and HbV: 4.21) and QM2 (PbV: 4.21, LbV: 4.00 and HbV: 4.17). ANOVA revealed no statistical differences among the three visualizations (PbV; LbV; HbV). These results indicated that the new visualization (HbV) was as useful as the other two visualizations. The average scores for visualization in QM3 were above 3.78, which indicate that those participants almost agree with the usefulness for reflection on what they did.

The correlation among participants and engagement measurement functions is presented in Table 2. This study results show that correlations between engagement measurement and student self-report are moderate for both PbA (r = .53) and IbA (r = .55).

In short, the results of the study show that:
- Writers felt they understand all the visualizations (agreement 4.06-4.26)
- Writers agreed with what the visualizations showed (agreement 4.0-4.21)
- Writers found the visualization useful to reflect on (agreement 3.78-3.96)
- No distinction perceived accuracy of the two engagement measurement algorithms was found.
- The height-based visualization is as useful as other visualization types.

### Table 1: Evaluation of Three Visualization Types

<table>
<thead>
<tr>
<th>Quality Measure</th>
<th>PbV</th>
<th>LbV</th>
<th>HbV</th>
</tr>
</thead>
<tbody>
<tr>
<td>QM1: Understand what the visualization is trying to convey.</td>
<td>M=4.26, SD=0.38, N=23</td>
<td>M=4.04, SD=0.59, N=23</td>
<td>M=4.21, SD=0.36, N=23</td>
</tr>
<tr>
<td>QM2: Agree with what the visualization is showing.</td>
<td>M=4.21, SD=0.54, N=23</td>
<td>M=4.00, SD=0.64, N=23</td>
<td>M=4.17, SD=0.33, N=23</td>
</tr>
<tr>
<td>QM3: Useful to reflect on what I did</td>
<td>M=3.78, SD=0.54, N=23</td>
<td>M=3.83, SD=0.42, N=23</td>
<td>M=3.96, SD=0.41, N=23</td>
</tr>
</tbody>
</table>

### Table 2: Correlation of Engagement Time

<table>
<thead>
<tr>
<th>Present Study N=23</th>
<th>PbA</th>
<th>IbA</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>PbA</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IbA</td>
<td>0.74</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>0.53</td>
<td>0.55</td>
<td>1</td>
</tr>
</tbody>
</table>
5.3 Impact of feedback on engagement

In order to evaluate the impact of feedback on participants' engagement, the average level of engagement among all participants were computed based on IBA (see Figure 9). Every 2 minutes the intensity of engagement of each participant was computed. According to Figure 9, almost all participants started the writing task with 100% engagement (continuously working on the writing) and their level of engagement reduced slightly as they approached submission time (t=30 minutes). 100% engagement (behavioral engagement) assumes that this student keeps writing the document all the time. If the time which the student spend on typing in writing less than the minimum threshold time, he or she is still fully engaged based on the IBA. The waiting time (from t=30 to t=40 minutes) is reflected well in the Figure 9. During this period participants received their feedback, read it and began to revise their draft. The participants were engaged again on the writing task after receiving and reading the feedback. Their level of engagement was slightly lower in the revising part from t=40 to t=60 minutes (M=51%, SD=17%) compared to the first half of the session (M=84%, SD=7%). We speculate that it might reflect the reduced concentration to the task by the participants in the second half of the experiment.

6 Conclusions

Behavioral engagement is difficult to track and measure without human intervention [10, 11] yet it has been identified as a frontier for affect-aware learning technologies[35]. The present study attempted to automatically capture the student behavior during a writing task by developing Tracer, a novel Learning Analytic system which uses Google API to collect the document's revisions, then analyses them and generates quantitative and visual measures of behavioral engagement over time. These visualizations successfully illustrated Johns' engagement measures [12], including intensity and consistency. The visualization evaluation results show that the average quality measure scores QM1 and QM2 were above 4. It indicated that writers agreed with what the visualizations conveyed and showed. In addition, the average scores for visualization in QM3 were above 3.78, which indicates that the participants almost agree that the visualization was useful to reflect on. In addition, the engagement measurement algorithms are useful since the correlation between engagement measurement and student self-report is moderate (r>.50).

One limitation of this study is that the impact of these visualizations on learning was not evaluated. Another limitation is that the performance of the high frequency revision sampling mechanism was not directly evaluated. Moreover, the current approach only considers the time that the writers were actively completing the activity. However, the findings from the current study suggest that visualising intensity could be useful to distinguish different student behaviors when approaching a writing task.

In future work, we will investigate when and how to use this tool to help students’ learning and how to evaluate its impact. Some researchers [27] have suggested to use the visualization derived from learning activity data of successful students to underperforming peers. One way to evaluate the impact of these visualizations is as follows: an online e-learning system will first display the personalized visualization and the visualization from a good student’s writing behavior to students at a certain stage of a writing activity. Then the system tracks when a student reads the visualization (a feedback intervention) and how many changes the student makes to the document when receiving the feedback. Checking how a student changes his/her behavior (i.e. making changes in the document) can be used to evaluate the impact of these visualizations.

Our system may eventually include estimations that consider additional dimensions of writing, including document content, length and topics included. In addition we have explored other modalities, such as facial
expressions and physiology signals that could be helpful in capturing emotional and cognitive engagement. This approach can also be adapted to measure engagement for collaborative writing since the Google Doc supports multiple people to synchronously work on the same document and keeps track of each person’s contribution on each revision.

ACKNOWLEDGEMENTS

The authors wish to thank the volunteers for participating in this study. Ming Liu and Rafael A. Calvo are supported by Young and Well Cooperative Research Centre. This work is partially supported by Fundamental Research Funds for the Central Universities under grant No. XDJK2014A002 and No. XDJK2014C141 and No. SWU114005 in China.

REFERENCES


E. Lane, “Clickers: can a simple technology increase student engagement in the classroom?,” in International Conference on Information Communication Technologies in Education, Corfu, Greece, 2009.


Techniques, 2011.


**A/Prof. Rafael A. Calvo**, PhD (2000) is Associate Professor at the University of Sydney’s and Co-Director of the Software Engineering Group that focuses on the design of systems that support well-being in areas of mental health, medicine and education. He has a PhD in Artificial Intelligence applied to automatic document classification and has also worked at Carnegie Mellon University, Universidad Nacional de Rosario, and as a consultant for projects worldwide. Rafael is author of over 150 publications in the areas of affective computing, learning systems and web engineering, recipient of five teaching awards. Rafael is lead Editor of the Oxford Handbook of Affective Computing and Associate Editor of IEEE Transactions on Affective Computing and of IEEE Transactions on Learning Technologies.

**Dr. Abelardo Pardo** is Associate Head of Teaching and Learning and Lecturer at the School of Electrical and Information Engineering, The University of Sydney. He has a PhD in Computer Science by the University of Colorado at Boulder. He is a co-director of the Software Engineering group and his research interest is in the application of technology to explore, understand and influence human behavior. He has participated in national and international projects funded by the Office for Teaching and Learning, National Science Foundation, and the European Union. Abelardo is author of more than 100 research publications in prestigious conferences and journals, member of the steering committee of the Society for Learning Analytics Research, associate editor of the Journal for Learning Analytics, and member of the editorial board of the IEEE Transactions on Learning Technology and the Journal of Social Media and Interactive Learning Environments.

**Prof. Andrew Martin** is an Educational Psychologist and Professor of Educational Psychology at the University of New South Wales (Sydney, Australia) specializing in student motivation, engagement, and achievement. He is also Honorary Research Fellow in the Department of Education at the University of Oxford, Honorary Professor in the Faculty of Education and Social Work at the University of Sydney, Fellow of the American Educational Research Association, and President Elect of the International Association of Applied Psychology’s Division 5 Educational, Instructional, and School Psychology.