The Use of Data Mining and Automated Social Networking Tools in Virtual Learning Environments to Improve Student Engagement in Higher Education

Stephen Smith, David Cobham, and Kevin Jacques

***Abstract*—Virtual learning environments (VLEs) form part of modern pedagogy in education; they contain student usage data that has the potential to inform and improve this pedagogy. The question this paper explores is how might the development of data mining and log analysis systems for the Moodle virtual learning environment improve students' course engagement? The paper proposes that a student will complete missed tasks sooner if their utilisation of the VLE is automatically tracked and electronic prompts are sent when VLE activities are missed. To explore and test the hypothesis a software tool, MooTwit was developed to contact students when they fell behind in their VLE study. To establish if student timely engagement improved the study used MooTwit with two groups of students over a period of 15 weeks, messaging one group only when they fell behind. Statistical analysis and comparisons were made between how quickly each group engaged with the missed items. Using MooTwit to track and contact students did influence the timeliness of their engagement with the VLE activities. Specifically, the results suggest by direct messaging a student to engage with missed material, they completed missed activities closer to required completion date. The findings within the thesis show that educational data mining has the potential to improve pedagogy in VLE linked education offering opportunities to increase timely engagement and to raise course designers' acceptance of data mining to improve the validity and quality of course evaluation.**

***Index Terms*—Virtual learning environments, student engagement, higher education, social networks.**

I. Introduction

The use of virtual learning environments (VLEs) to augment the learning opportunities through online courses, supplementary activities and resources for learners has been identified as being utilised in a high proportion of surveyed educational institutions within the UK [1], [2]. The adoption of such technologies has led to a wider range of tools to engage learners with topics relating to their areas of study. This increased reliance on resources that are used by students independently has created new challenges for the learner; for a learner to be successful in their studies they must engage with the VLE systems and make efficient use of the time they spend on the learning resources. These desirable outcomes very much link to the memorable quote "*Good habits formed at youth make all the difference*" made by the Greek philosopher Aristotle which inspired the intent of this study of encouraging learners to take ownership of their VLE learning to the point where they autonomously completed activities online.

There have been many studies focusing on VLE data mining [3]-[7], on learning analytics [8]-[11] and on data informed design [12] that identify the level of utilisation of the activities and resources within them. The common theme from these investigations is the success in generating the information from the data; the conclusions drawn from the studies consistently avoid establishing whether there was any positive impact from providing the information to either the course designer/deliverer or the learners themselves. Conclusions typically make statements such as "expert teacher knowledge for learning analytics can sometimes be outperformed by knowledge derived by data-mining algorithms" [5] and "Educators with no expertise in data mining can also apply their hands in these fields." [3] although Kaur does evaluate the success of the accuracy of the information in identifying slow learners, there is no evidence of the research having an impact on the students' performance. This study explores the development and impact of data mining an open-source Modular Object-Oriented Dynamic Learning Environment (Moodle) based virtual learning environment and automatically prompting learners who missed the deadline to access materials and activities through social networking and hypothesized that students receiving automated prompting would engage with missed VLE course materials in a timelier manner than unprompted students. The investigation was undertaken at Lincoln College a Further Education College in Lincolnshire, UK; the college provides further education and higher education courses for post-secondary and adult learners in the area.

II. Literature Review

Virtual learning environments are the most common systems used to the deliver learning materials and activities in UK further education establishments and considered a "…pervasive technology in higher education institutions…" [13]. The pedagogical underpinning for the use of VLEs is intended to enable a transition from teacher-led delivery to a more student-centred learning experience. Their adoption has been encouraged by the UK government and education funded Joint Information Systems Committee (JISC) who identify digital technologies such as VLEs as exciting opportunities to reconsider how we teach, engage with, or involve learners in different ways [14]. The use of VLEs

Manuscript received December 12, 2021; revised February 8, 2022.

Stephen Smith is with Lincoln College, UK (e-mail: research@stephensmith.me.uk).

David Cobham and Kevin Jacques are with the University of Lincoln, UK (e-mail: dcobham@lincoln.ac.uk, kjacques@lincoln.ac.uk).

doi: 10.18178/ijiet.2022.12.4.1614

present opportunities at the institutional, the teacher and the student level for improving and enhancing learning, hence there being a significant investment in the learning technology. Teachers who have little experience of the VLE technology as a teaching tool often miss the requirement to teach the learner how to use the VLE effectively, consequently the students "…find it difficult to engage with technology-based tasks as they may have little prior experience." [15]. More importantly students who do not engage with the learning may not be identified until they have fallen so far behind that they are hampered in their progress with coursework. This issue is a concern for the participants in this study, the Computer Studies section of Lincoln College who support students in their learning through using Moodle VLE courses to provide all learning materials and learning activities for the section.

A Moodle course is split into a structured set of activities. A typical Moodle course structure is shown in Fig. 1, where a Moodle course is made up of a sequence of topics (Moodle sections) containing a number of activities (Moodle modules); some of the activities are mandatory for a student to complete the course at a minimum level and are expected to be completed in a timely manner.

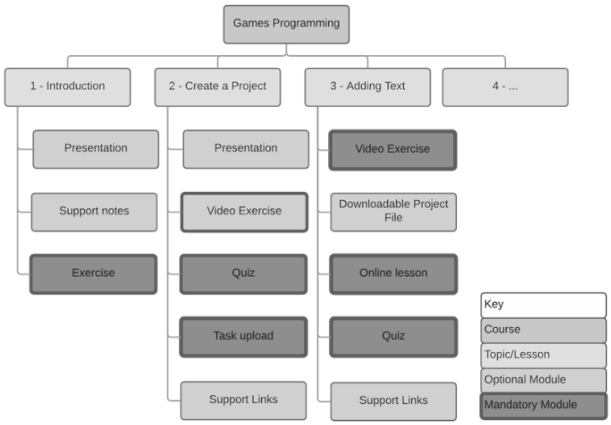

Fig. 1. A typical section of a Moodle course structure.

The course presents itself to the student through a web interface, that allows them to navigate the sections and engage with modules. The look and feel of the interface can be modified through the application of themes, Fig. 2 A Sample section of a Moodle course presented to a computing student at Lincoln College is shown. There are four modules within the section the student should engage with:

- Introduction to XNA presentation (Pop-up URL module).
- Exercise 01 Creating an XNA project video (Page module).
- Adding text to an XNA Screen presentation (Pop-up URL module).
- Exercise 02 XNA writing to the screen (Lesson module).

'Exercise 02 XNA Writing to the Screen' is an example of a Moodle lesson module which would present itself to the student as a navigable lesson with quiz elements.

Within Lincoln College's computing section, the expectation of all tutors is that students are all fully engaged with their courses on the VLE, and it has been observed that in many instances students have struggled to keep track of online exercises and tasks set. This situation is confirmed in another study of VLE engagement that concluded "...although some students engaged extensively with the online activities, many did not or did so less than the course team and the teachers had hoped." [16]. A lack of engagement with the VLE in this scenario as identified by Morgan, cited in [17] can have an impact on the level of success of a student. Kuh *et al.* identified two key components for a student to be successful, "…the amount of time and effort the student puts into their studies…" and "…the ways the institution allocated resources and organizes learning opportunities and services to induce students to participate…" [18]. In a study carried out by the University of Wolverhampton students were encouraged to participate in VLE learning activities using the Wolverhampton Online Learning Environment (WOLF); the analysis of the study recommended "…monitoring of regular habits in using WOLF, including the activities" [11] to ensure adequate engagement. Mazza & Milani observed that, "Educational research literature also shows how monitoring student learning is a crucial component of high-quality education." [19], but for a tutor to check a student's engagement with a Moodle course they are required to manually navigate a HTML forms-based query system and analyse the resulting report to identify what the student has and has not accessed [20]. The process is labour intensive; regular application to a full student cohort would create significant workload for a tutor. With software automation this is an avoidable scenario. It has been identified that Moodle "…collects a large amount of data on student interactions within it, including content, assessments, and communication…" [6] that could be data mined.

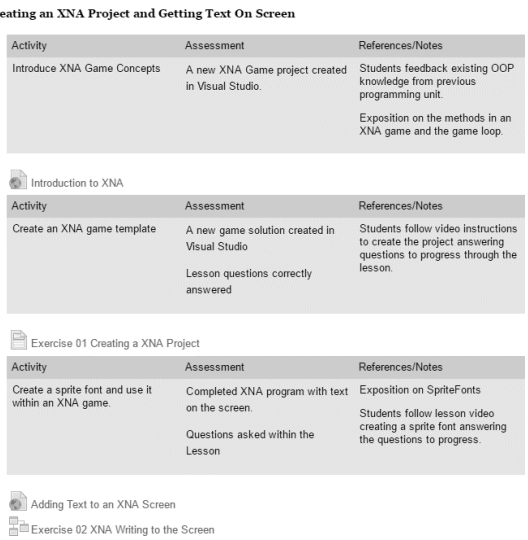

Fig. 2. A Sample section of a Moodle course.

Fig. 3. A sample Moodle lesson module.

Solutions have already been attempted to improve the quality of the reporting provided to tutors. Previous research and systems developments focused on the use of log analysis to produce a "learning analytics dashboard" [21], charts and graphs alongside materials in Moodle [19] or "...provides aggregated and useful statistical reports." [10]; all require the direct engagement of the designer and deliverer to interpret and action the results; typically, non-engagement of a student is resolved and identified by the deliverer, from this it expected that they will be able to identify weak spots in the course and amend the design for the next delivery session. The previous studies reporting systems have a variable level of effectiveness in terms of their ease of use, the expected technical level of the users of the systems and a user's ability to interpret the information presented. The software developments within this study aim to provide better tools to track and prompt disengaged learners and simplify data mining of resource utilisation to improve the course quality, both having the end goal of improving course engagement.

None of the previous research has covered the use of automated stimulus control as a mechanism for improving student engagement with a VLE. Stimulus control is defined by Bloh as "…a change in operant behaviour that occurs when a particular type of stimulus is presented" [22]. The findings of Miltenberger on prompting and transferring of stimulus control indicated that through multiple prompts behaviour can be changed [23]. This study proposes the use of social networking as the stimulus based on the observation that "these applications already gained high popularity among students and are suitable to be used to engage the students…" [7]. This is supported by existing evidence of social networking being used in education, for example Junco *et al*. found that "…using Twitter in educationally relevant ways can increase student engagement and improve grades" [24] and Twitter use significantly increasing student use of valid Internet sources in research as identified by Halpin, [25]. Research undertaken by Knight & Kaye into students preferred educational use of Twitter also identified that, "The three most highly ranked suggestions were the use of Twitter to provide information on practical issues, the posting of course-related reminders and the post of assignment submission deadlines" [26]. These studies provide indicators that the use of social networking could yield positive results for this study.

In summary, the goals of this research are to establish whether it is possible to improve the timeliness of student engagement with VLE activities/resources. This is achieved through the development of a software artefact MooTwit – a Moodle plug-in enabling checking student engagement with the VLE and automating the process of prompting the student regarding their lack of timely engagement and examining the change in the study participants' behaviours and attitudes because of the utilisation of the software. The system will:

- Log analysis to identify a student's level of engagement.
- Twitter and Email messaging directing students who are late accessing the learning materials on the VLE and congratulating students who complete work on time.
- Messages to course deliverers regarding lack of individual student engagement.

This development, use and the accompanying analysis of the results of the plug-ins use provides an insight into how students engagement with the course changes and offers guidance for other researchers who would like to develop VLE plug-ins.

III. Methods

When considering the research methodology there was a need to consider collecting of data that would allow for the evaluation of the success of the system. A quantitative approach was selected where the research makes use of data from the Moodle database logs of the students' VLE accesses; this aligns with Neall's quantitative criteria including, "Data is in the form of numbers and statistics" and the research is looking for specific changes in engagement meeting the criteria "Researcher knows clearly in advance what he/she is looking for" [27].

The possibility of bias in the research was considered; while "bias can never be completely eliminated" [28] the methodology endeavoured to avoid bias as follows:

1) By automating the process of logging of virtual learning accesses any data collection activity is invisible to the participants, giving a lower chance of participant bias.
2) Bias due to experimenter behaviour has been reduced by using automated social media messaging generated by the engagement system ensuring all participants receive the same feedback from the system at the same time in the same format.

The use of log analysis removed the opportunity for the learner to be aware of what they are being tracked on; obfuscation of the tracking prevents them from changing their behaviour until the system informs them that they missed the milestone for completing the task. The removal of the tutor from the process of tracking and prompting of the students limits the influence they have upon the students when they miss the work, lowering the impact on how quickly the student might complete the missed work.

The system was developed as a proof of concept that a plug-in linked to Moodle and Twitter could be used to track and improve timely engagement with important course resources on the VLE. As a proof of concept, the plug-in is scalable in line with the database and server systems that support the VLE; its restriction is the extent that social media limits the use of automated direct messaging.

A. Summary of Plug-in Development

Development was limited to the Moodle VLE based on the skill set of the users involved in the study and the system being used operationally within Lincoln College. The main development language for the engagement system was PHP: Hypertext Pre-processor (PHP) a HTML-embedded scripting language. The rationale for this was based on it allowing "web developers to write dynamically generated pages quickly" [29] and its compatibility with the Moodle VLE which is also written in PHP. This aided the development through the ability to call the existing Moodle log API, [20]. To enable communication of the level of engagement to a student required a link to a current social networking messaging infrastructure for direct messaging; Twitter was selected as the most appropriate system for the experimental

direct messaging. A supporting factor for its choice was that Lincoln College use Twitter with students on a regular basis and its API is open to developers to use freely; however, at the time of evaluation there was a restriction of "250 direct messages per day" [30] per user, which does limit the scalability of the product. The hosting service selected for this study offers a built-in PHP mail function that facilitates the sending e-mails directly from a PHP script [31] that is to be used for delivering a summary of course elements missed each week to both the student and the course coordinator.

To integrate the required functionality into the Moodle VLE a Moodle plug-in was selected as the target for development in preference to the construction of a discrete website to perform the task. The Moodle VLE has the capability to extend itself through the use of a plug-in-based system which is the most straightforward approach to adding new functionality to Moodle (Moodle Docs 2016b). The use of this system also ensured that access to the data relating to the student was afforded the same level of security as the VLE, conforming with the Technical element of the learning analytics "DELICATE checklist" [32].

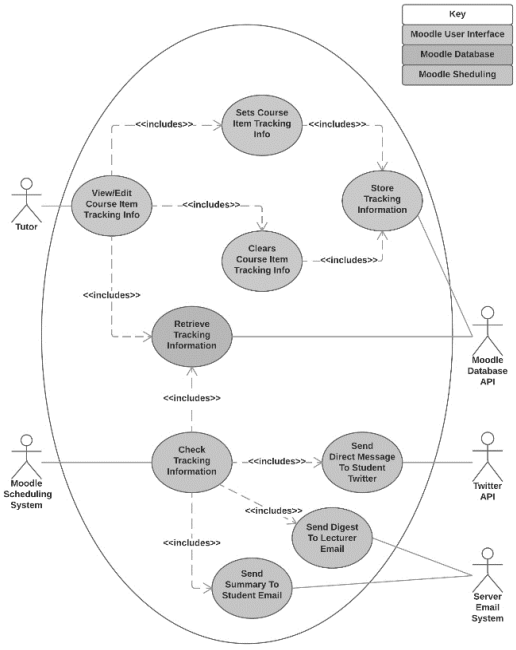

Fig. 4. Use case diagram student engagement system.

The plug-in type selected was based on the Moodle report plug-in type designed to provide "…useful views of data in a Moodle site" [33]. The process of developing the plug-in requires the construction of a number of server-side scripts written in the PHP scripting language to enable the installation, upgrade and running of the plug-in. The software design of the plug-in requires the use of both procedural design for the HTML/PHP page display and automated CRON (Command Run ON) functions; in line with the Moodle design principles, object-oriented development was used for the user interface elements. The plug-in also requires the design and implementation of tables in the Moodle database to allow data persistence and constructing SQL (Structured Query Language) queries to examine the existing Moodle database tables to retrieve course and student access data. A UML use case model [34] was created as shown in Fig. 4 to establish the key functional elements of the system prior to development.

The Moodle plug-in presents itself to the tutor as an embedded form within the virtual learning environment, allowing the tutor to flag individual parts of a lesson as required learning and specify completion dates.

A tutor's submission of the form within the VLE stores the tracking information to the Moodle plug-ins database tables for use in its background CRON function, providing the core automation for the plug-in to perform the following tasks:

1) Retrieving monitored tasks from the database.
2) Checking student completion of monitored tasks.
3) Contacting students to inform them of incomplete tasks.
4) Creating and emailing to registered staff a summary of student progress.

B. Summary of Study Operation

To establish if student engagement can be improved by log analysis and direct message prompting the Moodle plug-in shown in Fig. 5 MooTwit engagement plug-in was developed to be used by the course subject tutor. This allowed the tutor to select either full sections of a course to be completed by a specified date or individual activities. Students who failed to complete an activity by the deadline were digitally messaged once per week for two weeks to prompt them to engage with it.

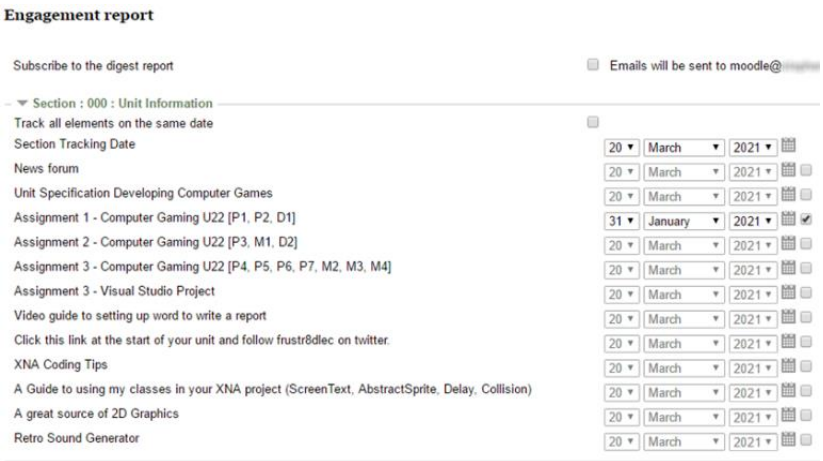

Fig. 5. MooTwit engagement plug-in.

The information required to identify success or failure in completing activities was gathered from the VLE log files by the plug-in. Students were then prompted with an automated Twitter direct message identifying they had missed work in the last two weeks Fig. 6 Twitter direct message (Failed to complete work), backed up with a more detailed email identifying the areas of study missed shown in Fig. 7.

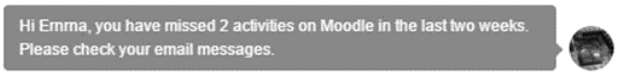

Fig. 6. Twitter direct message (failed to complete work).

To: [Student Email]
From: [Lecturer Email]

Subject: Engagement Report

Hi Erna,

You seem to have missed some work on Moodle in the last two weeks. To catch up you need to complete the following:

Unit 22 Developing Computer Games in section 0 Unit Information, Assignment 1 - Computer Gaming U22 [P1, P2, D1] should have been completed by: 31 January 2021

Unit 22 Developing Computer Games in section 2 Assignment Computer Gaming, Survey Question - I understand what I am required to do to complete assignment 1 should have been completed by: 31 January 2021

Regards

[Lecturer Name]

Fig. 7. Sample missed activity email.

If the student had completed all the work a message congratulating them was sent instead Fig. 8 Twitter direct message (Completed all work).

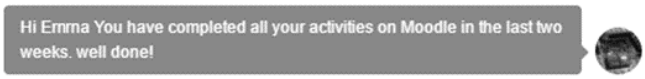

Fig. 8. Twitter direct message (Completed all work).

The study was performed using two separate classes of students, both were second year groups of a level 3 BTEC Extended Diploma in IT course who had joined with similar entry criteria and studied the same subjects with the same tutors prior to the intervention commencing. One class formed the control group, the other the experimental group, both were informed of the intervention and consented to take part but were unaware of whether they were the control group or experimental group, neither group was aware of the others participation. To reduce the possibility that one group might influence the other, communications about the study between the two groups was limited by timetabling their courses on different days of the week.

The study took place over 15 weeks with 14 students in the control group and 15 in the experimental group. The control group originally had 15 members, but one member was taken ill in the first week of the study and was unable to contribute.

In the first week of the study both groups were prompted to provide their Twitter name. For the study the control group would be opted out of the engagement system and their Twitter and email accounts would not receive any communications. Over the course of the 15 weeks 57 monitored deadlines were set.

The evaluation of the success of the system was defined by the hypothesis that the experimental group will access missed material sooner than the control group because of being prompted by the system. The null hypothesis was that the experimental group will not access the learning activities any earlier than the control group after being prompted. Evaluation of the engagement system involved the comparison of log data from the control group with the experimental group. To complete this process, the data from the log table in Moodle was exported to a single table which was then split into relational tables as shown in Fig.9 Entity relationship diagram exported log data.

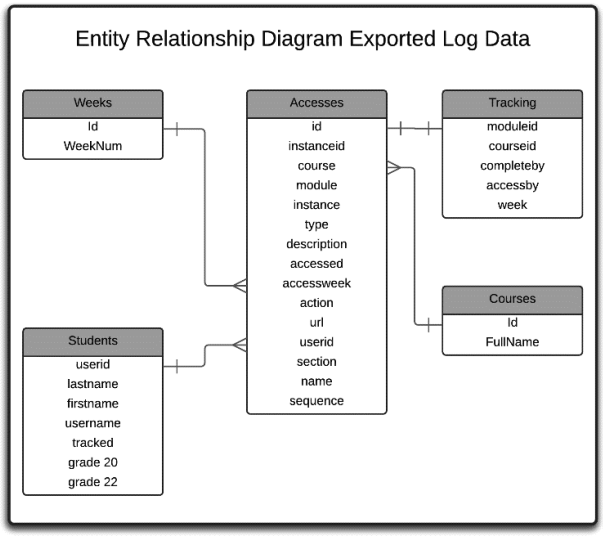

Fig. 9. Entity relationship diagram exported log data.

To enable the tracking of the student behaviour in a consistent manner the tracking dates were resolved into weeks from the start of the course; access measurements were then based around the week of access to ease the classification of timely access to the learning activities. Access results were based on the first access a student made to a learning activity and were categorised as:

Positive accesses	took place by the deadline or within the 2-week window of prompting (student is on track).
Late access	took place after the 2-week window of prompting (student has fallen behind).
Missed access	the student failed to access the activity within the limits of the course.

For each group, the accesses were tabulated as shown in Table I Sample record of accesses per week, the total number of weekly accesses after the expected week of access being recorded.

The evaluation looked at the changes in student access to the VLE on both a group access level measuring the changes before and after the system; and further inspection at regular periods during the research to gain further insights into variations in engagement over time. Conclusions drawn from the evaluation will identified the successes and/or failures of the system in improving engagement and provide a model for others to replicate with other student groups.

IV. Analysis

Student engagement was measured by their timeliness of access to the resources on the VLE; a student was determined to be appropriately engaged with the material if they accessed it either before the expected access date or within a window of two weeks after the date; if a student failed to access the material after that date they were considered not to be appropriately engaged with the materials.

Prior to interpreting the quantitative data from the log files statistical tests were undertaken to ensure the results were statistically significant. IBM SPSS software was utilised to extract relevant results and an additional calculation for identifying the effect size.

Performing the statistical tests required some pre-processing of the captured data using the criteria identified below.

- Positive accesses ≤ 2 weeks of expected access date.
- Negative accesses > 2 weeks of expected access date or no access made.

The groups were split into the categories.

- Tracked – the experimental group who were prompted to engage.
- Untracked – the control group.

TABLE I: Sample Record Of Accesses per Week

	Activity Description	Upload your controller information here.	Example details of how to describe a game technology
	Access by Week	2	3
Positive Access Counts	On Time/Before Prompt	2	
	After First Prompt	10	7
	After Second Prompt	1	1
Negative Access Counts	Over 2 Weeks Late		
	Over 3 Weeks Late		2
	Over 4 Weeks Late		1
	Over 5 Weeks Late		
	Positive Accesses Total	13	8
	Negative Accesses Total	0	3
	Total Accesses	13	11
	Max Possible Access	14	14

A. Independent Samples t-Test

An independent-samples t-test was conducted to compare the timeliness of VLE accesses in tracked (electronically prompted to engage with missed activities) and untracked conditions. There was a significant difference in the scores for prompted (M = 0.54, SD = 0.50) and unprompted (M = 0.33, SD = 0.47) conditions; t (1738) = 9.24, p = 0.0002. Full details of the t-test results are shown in Table II.

These results suggest that electronic tracking and prompting of students who miss activities does influence the timeliness of their engagement with the VLE activities. Specifically, the results suggest that when students receive a Twitter direct message to engage with missed material, they complete missed activities closer to required completion date.

A small to medium (0.43) effect was identified when the Cohens d calculation [35] was applied (1) to the results.

$$\text{Effect Size Cohens d} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}} = 0.43 \qquad (1)$$

The experiment's hypothesis indicated that students who were tracked would engage more than non-tracked students, given that T value is greater than critical value 1.96 obtained from table B2 [36] the null hypothesis that both groups would be equally engaged cannot be accepted.

TABLE II: Independent Samples t-Test

	On Time Access	
Tracked	Yes	No
N	900	840
Mean (X)	0.54	0.33
Std. Deviation (σ)	0.5	0.47
S.E. Mean	0.02	0.02

	On Time Access	
	Equal variances assumed	Equal variances not assumed
F	94.6	
Sig.	0	
t	9.23	9.24
df	1738	1737.77
Sig. (2-tailed)	0	0
Mean Difference	0.21	0.21
Std. Error Difference	0.02	0.02
Lower	0.17	0.17
Upper	0.26	0.26

B. Comparison between Accesses

From the results of the t-test and Cohens d it was identified that the hypothesis for students being prompted accessing the materials earlier was most likely and that the effect size was close to being classified as medium.

From Fig. 10 Graph of student access profiles it can be seen that the positive accesses for the experimental group was 22% higher and they completed 19% more activities than the control group. The experimental group also performed 3% better than the control group on late accesses.

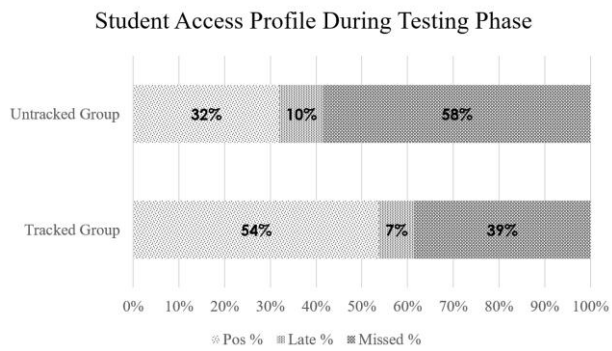

Fig. 10. Graph of student access profiles.

The high levels of late and missed activities by the untracked group provides a strong argument for the need for tracking of student engagement given that the tracked activities were flagged as mandatory elements of the courses.

C. Positive Access Trend Comparison

Over the duration of the experiment, the prompting was applied multiple times to the experimental group to identify if the process would show a positive or negative cumulative effect on them in comparison with the control group. Miltenberger's work indicated that through multiple prompts the engagement should increase [23]. From Fig. 11 Engagement trends during study, it has been observed that the experimental group showed a positive trend in the mean timeliness of access of 16% from the start of the experiment indicating that repeated prompting does have a positive cumulative effect over time. In comparison, the control groups' improvement over time has not increased and is 34% lower than the experimental group by the end of the study.

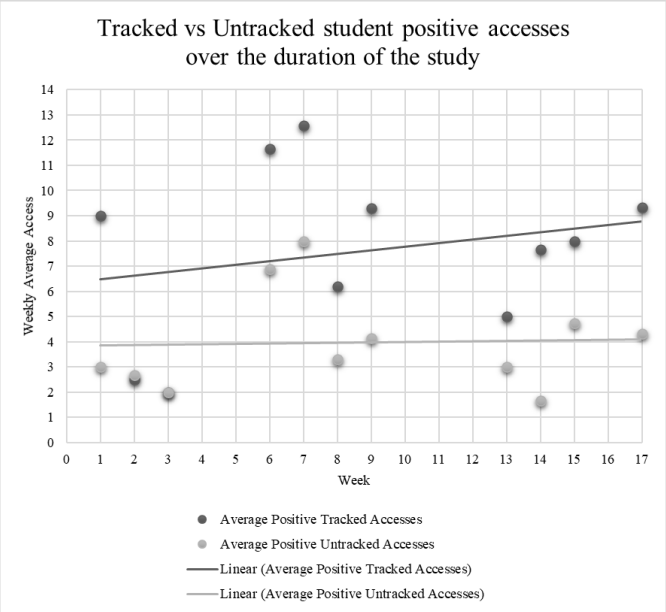

Fig. 11. Engagement trends during study.

When evaluating the results, it was noted that the untracked group showed an initial lower level of performance and it remained so throughout the study compared to the tracked group; this does not however, relate to how the behaviour of students was altered by prompting as the study compares the improvement from each groups start point. In Fig. 11 the trend of the prompted participants showed continuing improvement from the participants starting point compared to the static results of unprompted students giving a good indication that there was an increase in engagement in the experimental group because of the messages being sent to them. The data points in weeks 6 and 7 on the graph indicate a higher level of completions by both groups, this can be explained by the delivery within those weeks being lighter around the half-term break within the 15 weeks of study, allowing a greater opportunity both in class and during the week break for some tasks to be completed.

D. Analysis Summary

Overall, the data analysis of the results of the study showed a positive increase in performance by the experimental group. The results of the study suggest that the use of a software solution to create an electronic message as a stimulus to produce a change in operant behaviour was successful within the parameters of the experiment effectively rejecting the null hypothesis was that the experimental group would not access the learning activities any earlier than the control group after being prompted.

When reflecting on the success of the research there is a need to review the limitations of the study including the use of a small sample of computing students within one institution which may not reflect the range of learners with the larger academic community. The duration of 15 weeks for the study may not have allowed time to establish if the performance gains made by the experimental group could be further improved by the continued use of the system or whether the learners might become disillusioned with system and cease to respond to prompts.

The use of extracted quantitative data from the Moodle database gave clear empirical evidence to establish the student's activity and to identify changes in behaviour and was an effective research methodology in this context simplifying data collection and processing.

The extraction of existing data from Moodle database using a plug-in was effective within the scope of the experiment and validates the existing research by Mazza *et al.* [9] and Zhang & Almeroth [10] that focused on the extraction and reporting from the system.

It is important to note that the plug-in development process was hampered somewhat by the limited documentation relating to the structure of the data held within the system and omissions in the open source Moodle documentation. Despite the development issues and that Moodle is an open source project, the plug-in system proved to be completely stable and performed reliably for full duration of the research suggesting it should be considered a viable system for conducting research on student's utilisation of the Moodle VLE.

Early in the research the literature review identified that while Mohamad & Tasir [7] proposed that the use of social networking was a possible communications mechanism, out of all the social networking providers only Twitter provided an API to allow for direct messaging and this had limitations on the number of messages that could be sent in 24 hours.

Although Twitter automatically gave a direct messaging system on a number of destination platforms including all mobile phones, tablets and PCs it was noted that a small number of learners did not use Twitter and had to sign up for accounts.

Given the limited messaging API provision by social networking providers and the range of providers that may or may not be used by learners, social networking may not be the best solution; messaging applications may offer a better system of communication, Facebook's WhatsApp messenger for example offers a messaging solution using a non-rate limited paid for business API to send direct messages to the App [37].

It is clear this investigation in isolation can only provide a small contribution to larger fields of study relating to the motivation of students to learn autonomously, what it does

provide is an indication that students respond positively to automated prompting. It is hoped that the results will promote discussion and further study into other automated mechanisms that assist with learner engagement using prompting mechanisms. The tool developed for the investigation can be used as is in Moodle installations and provides a framework for future Moodle plug-in developers both for research and for the development of plug-ins for general use. For course designers and tutors the plug-in provides a stable system for contacting small groups of students when they do not engage with important learning activities; it also provides a methodology for identifying if there is measurable improvement in engagement for their students. Should the plug-in be used in other academic institutions the system should at least provide helpful prompts for students to enable them to easily identify missed work before falling too far behind in their studies.

V. Conclusion

All studies previously undertaken have focused on the methods of analysing and displaying information but have not investigated if this could be used as a positive effect on the student. The primary goal of the MooTwit development was to extend the data mining process into a system that improved student engagement with VLE courses by analysing their accesses and automatically prompting them via social networking backed up with email when they were falling behind in accessing activities on the VLE.

The hypothesis was that the experimental group (students receiving automated prompting) would access missed VLE course materials sooner than the control group. The results of the study support this hypothesis suggesting that electronic tracking and prompting of students who miss activities does have an a small to medium effect on the timeliness of their engagement with the VLE activities. Specifically, the results suggest that when students receive a Twitter direct message to engage with missed material, they complete missed activities closer to required completion date compared to the unprompted students. In the context of the initial aim of the research of improving engagement of students with the activities, clearly there has been a demonstrable change in behaviour by students in the experimental group. There is a good indication that the adoption of the system would be beneficial, increasing student utilisation of course resources at the appropriate point in their studies, enhancing the quality of their learning experience. There are some limitations to the extent of this success in that these results may not be reflected in the context of the wider learner community; this experiment focused on a specific group of students; the success shown here may not yield the same result with different learner demographic. This limitation provides scope for extending the investigation into a larger domain to establish if the findings can be completely generalised by varying such parameters such as the level of technological ability, initial engagement level of the students with the VLE, the age of the participants or course subject studied.

In the preliminary stages of reviewing the results of the research there was a concern that the use of prompting in the system may have demotivated the students or that the effect of prompting would be less effective over time. The initial statistical data analysis addressed demotivation in terms of the statistical outcome of prompting having a positive effect; it did not however consider the use of the prompting over time. The students may have been receptive to the system at the start of the experiment, but less so by the end of their studies. The use of a t-test and Cohens D effect level did not provide a temporal measure to allow this level of analysis. The study addressed this limitation by analysing the student behaviour over the duration of the experiment identifying that the experimental group showed a positive trend in the mean timeliness of access to course resources over the 15 weeks of the experiment indicating that repeated prompting does have a positive cumulative effect over time. This result in combination with the findings of Miltenberger [23] on the use of prompting and transfer of stimulus control would suggest the automated prompting has a similar effect to prompting by a course deliverer and that there is a transfer of stimulus of control over time. As with most studies of this scale there is opportunity to extend this investigation both in terms of the refinement of the methods used to prompt the participants and to further study participant behaviour (good habits) post experiment, given that the intent of the process was to improve engagement it is especially important to confirm if the effects were temporary. A longer study may also highlight if there would be an element of apathy in responding to the prompts as they became the accepted norm.

Overall, this investigation into improving student engagement provides evidence to support the case for data mining of VLE data stores to aid in establishing good study habits in learners. While social networking as a form of communication was used successfully as part of this small study, the restrictions in many social networking infrastructures limit the implementation and scalability of the process in its current form. There are also concerns over the volatility of the popularity of social networks; while at the time of writing Twitter still has popularity at the time of writing, Instagram for example has a growing user base. On the basis of the varying level of adoption of different social networking sites is it suggested that future areas for study could include the use of other low cost and low latency communication systems such as Skype or Google Firebase Cloud Messaging for communication of prompts to the learners as an alternative to social networking effectively eliminating the issue of which social network to use and providing a dedicated path of communication between the system and the student.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

The content of this paper is based on research carried out by Stephen Smith for the award of MSc by Research under the supervision of Dr David Cobham and Dr Kevin Jacques. The original thesis can be found at reference [38] below. All authors contributed to the writing of the paper and all authors have approved the final version.

REFERENCES

[1] UCISA, "Survey of technology enhanced learning for higher education in the UK," 2018.

[2] Ofsted & The Office for Standards in Education, "Virtual learning environments: An evaluation of their development in a sample of educational settings," *Alexandra House*, 2009, pp. 1–28.

[3] P. Kaur, M. Singh, and G. Josan, "Classification and prediction based data mining algorithms to predict slow learners in education sector," *Procedia Computer Science*, vol. 57, 2015, pp. 500–508.

[4] S. K. Mohamad and Z. Tasir, "Educational data mining: A review," in *Proc. Social and Behavioral Sciences*, vol. 97, 2013, pp. 320–324.

[5] C. Romero, C. Castro, and S. Ventura, "A Moodle block for selecting, visualizing and mining students usage data," presented at 6th International Conference on Educational Data Mining, Memphis, TN, USA, 2013.

[6] R. Mazza *et al*., "MOCLog — Monitoring online courses with log data," in *Proc. 1st Moodle Research Conference*, 2012, pp. 132–139.

[7] A. Ortigosa and R. M. Carro, "The continuous empirical evaluation approach: evaluating adaptive web-based courses," in *Proc. 9th International Conference on User Modelling*, Berlin, Heidelberg: Springer-Verlag, 2003, pp. 163–167.

[8] D. Y. T. Liu *et al*., "Knowledge acquisition for learning analytics: comparing teacher-derived, algorithm-derived, and hybrid models in the moodle engagement analytics plugin," *Pacific Rim Knowledge Acquisition Workshop*, 2016, pp. 183–197.

[9] D. Yen *et al*., "An enhanced learning analytics plugin for Moodle: student engagement and personalised intervention," *Globally Connected, Digitally Enabled. Proceedings of Ascilite 2015*, pp. 180–189.

[10] H. Zhang and K. Almeroth, "Moodog: Tracking student activity in online course management systems," *Journal of Interactive Learning Research*, vol. 21, no. 3, 2010, pp. 407–429.

[11] C. Dale, "A wolf in sheeps' clothing? An analysis or student engagement with virtual learning environments," *The Journal of Hospitality Leisure Sport and Tourism*, vol. 6, no. 2, 2007, pp. 100–108.

[12] K. Currie, "Scalable and repeatable development with a personalised touch," *Moodle Moot Ireland UK*, 2016.

[13] M. Weller, *Virtual Learning Environments: Using, Choosing and Developing Your VLE*, Taylor & Francis, 2007.

[14] JISC, "Technology and tools for online learning," 2016.

[15] G. Hughes, *Learning-to-Learn Online: Fostering Student Engagement with Online Pedagogies*, Taylor & Francis, 2005.

[16] R. Hampel and C. Pleines, "Fostering student interaction and engagement in a virtual learning environment," *CALICO Journal*, 2013, pp. 342–370.

[17] A. Maltby and S. Mackie, "Virtual learning environments — Help or hindrance for the "disengaged" student?" *ALT-J, Research in Learning Technology*, 2009, pp. 49–62.

[18] G. D. Kuh *et al*., *Student Success in College: Creating Conditions That Matter*, Wiley, 2011.

[19] R. Mazza and C. Milani, "GISMO: A graphical interactive student monitoring tool for course management systems," in *Proc. T.E.L.'04 Technology Enhanced Learning '04 International Conference*, Milan, 2004, pp. 18–19.

[20] Moodle Docs, "Logs – MoodleDocs," 2013.

[21] P. Dawson, "A learning analytics block to identify students at risk of disengaging with their studies – $50k," *NetSpot Innovation Fund*, 2014.

[22] C. Bloh, "Assessing transfer of stimulus control procedures across learners with autism," *The Analysis of Verbal Behavior*, vol. 24, 2008, pp. 87–101

[23] R. Miltenberger, "Behavior modification: Principles and procedures," *Cengage Learning*, 2011.

[24] R. Junco, G. Heiberger, and E. Loken, "The effect of Twitter on college student engagement and grades," *Journal of Computer Assisted Learning*, vol. 27, no. 2, 2011, pp. 119–132.

[25] B. P. A. Halpin, *Using Twitter in a Nonscience Major*, vol. 45, no. 6, 2016, pp. 71–78.

[26] C. G. Knight and L. K. Kaye, ""To tweet or not to tweet?" A comparison of academics' and students' usage of Twitter in academic contexts," *Innovations in Education and Teaching International*, vol. 3297, April, 2015, pp. 1–11.

[27] J. Neill, "Qualitative versus quantitative research: Key points in a classic debate," 2016.

[28] J. Lazar, J. H. J Feng, and H. Hochheiser, *Research Methods in Human-Computer Interaction*, Wiley 2010.

[29] The PHP Group, "PHP: General information," 2017.

[30] Twitter. Twitter help center | Twitter limits (API, updates, and following). [Online]. Available: https://support.twitter.com/articles/15364-about-twitter-limits-update-api-dm-and-following

[31] 1&1 Internet Inc. Send e-mails using PHP. [Online]. Available: https://help.1and1.com/hosting-c37630/scripts-and-programming-languages-c85099/php-c37728/send-e-mails-using-php-a657019.html

[32] H. Drachsler and W. Greller, "Privacy and analytics — It's a DELICATE issue a checklist for trusted learning analytics," in *Proc. 6th Conference on Learning Analytics and Knowledge*, 2016, p. 10.

[33] Moodle Docs. Plugin types — MoodleDocs. [Online]. Available: https://docs.moodle.org/dev/Plugin_types

[34] G. Curtis and D. Cobham, *Business Information Systems: Analysis, Design & Practice*, 6th edition, Pearson Education M.U.A. 2008.

[35] J. Cohen, *Quantitative Methods in Psychology*, vol. 112, no. 1, 1992, pp. 155–159.

[36] N. Salkind, *Statistics for People Who Hate Statistics*, SAGE Publications Ltd. 2014, p. 372.

[37] P. Olson, "WhatsApp is finally inviting businesses onto its massive network this year," *Forbes*, 2016.

[38] S. Smith, "Thesis for the award of MSc by research," University of Lincoln Library.

Copyright © 2022 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).

Stephen Smith has been teaching, researching and developing digital education systems for 25 years. His interests include computing, education and interactive learning technology systems since graduating in computer science and then completing his MSc in computer science at Lincoln University.

Throughout his career he has been a computer hardware developer, independent software developer, lectured in computer science and is the Departmental Lead for Computer Science at Lincoln College.

Mr. Smith is a Certified Microsoft Innovative Educator.

David Cobham has been teaching, researching and providing consultancy in I.T. and information systems for over 35 years.

His interests span business, computing and education having originally graduated in Economics and then completed postgraduate programmes in Mathematics at Kings College London and Computer Science at City, University of London, then a PhD in Information Systems and Constructive Alignment in Higher Education from the University of Lincoln.

Throughout his career he has been a departmental leader holding the posts of Head of School/Department of Computer Science, and Dean of Faculty. Amongst his publications is the core university text "Business Information Systems: Analysis Design and Practice", which has run to six editions, has been reprinted and translated multiple times and has over 500 citations recorded on Google Scholar.

Dr. Cobham is a Principal Fellow of the UK Higher Education Academy and is a Fellow of the British Computer Society and a Chartered IT Professional.

Kevin Jacques has worked at the University of Lincoln since 1998 and has a keen interest in information systems and the use of them to promote pedagogic excellence. He graduated from the University of Lincolnshire and Humberside with a degree in computer information systems and completed his doctoral studies in educational policy and values at the University of Hull.

His career covers a wide range of academic roles from programme leader to deputy head of School and to his current role as director of Academic Quality and Standards.

Dr. Jacques is a senior fellow of the UK Higher Education Academy, a fellow of the Institute of Analysts and programmers and a member of the British Computer Society. He was the inaugural winner of the Vice Chancellor's Teaching Award for Academic Excellence in 2015 and BTEC UK Student of the Year in 1997.