Defeating Hawthorne in tech-enabled education: Passive observation of student behaviour with a remote laboratory

Joshua Burridge[a], David Lowe[a], Keith Willey[a], and Judy Kay[a].
Faculty of Engineering, University of Sydney
Corresponding Author Email: joshua.burridge@sydney.edu.au

Introduction

Educational laboratories have the potential to be a highly interactive form of teaching and learning activity involving actions between students and equipment. The constrained nature of the equipment means that these actions tend to be both more measurable and more able to be classified (e.g. “started the pump”, “adjusted the resistance to 5 ohms”) than is the case for the interactions that occur in less structured educational environments or activities (i.e. conversation or writing). A consequence of this is that we have a potential opportunity, through collecting and analysing this data on students’ laboratory interactions, to gain insights into the relationship between the design of the laboratory experience and the student learning behaviours. This potentially may inform enhancements in the way in which we design these learning activities.

Researchers might also obtain insights through more invasive forms of observation such as in-person visual observation, camera recordings, or survey questions – however, as famously indicated in the Hawthorne experiment (Roethlisberger & Dickson, 1939), these forms of invasive observation may have unintended effects on the outcomes they observe. The method we use in this work involves a form of “passive observation” where we simply use data the system needs to capture. We avoid making students feel that they are being watched. So it ‘sidesteps’ the potential problem of the act of observation influencing student behaviour and thus research results, and avoids the issue of selection bias caused by participants opting out of active observation or querying. Our method needs to operate within the constraints of such passive observation and associated limits to the data richness.

This paper reports on an empirical study of a truss laboratory used in civil engineering education. This laboratory is delivered remotely; this provides valuable data collection without invasive ‘active’ questioning or observing of students directly – indeed, the data used in this paper was collected without prior intent for research because it is needed for the regular operation of the existing remote laboratory. This data has revealed patterns of equipment use, which has subsequently informed the design of improvements which resulted in large practical benefits from increases in utilisation rates (up to a potential yield of 1500%). From a pedagogic perspective, we have identified distinct student behaviours that may potentially be used for real time learner modelling and adaptive laboratory activities. We demonstrate the usefulness of this data and highlight ways that others could also use similar data when it is available, and to inform design of data recording to provide similar valuable insights without the burden of active research participation.

The Laboratory System

The laboratory apparatus under study is a static truss which is remotely controlled via the internet. Students are able to log in and control this truss by applying a variable load at a variable angle to a single point on this structure. The equipment includes strain gauges on each member reporting data to the web interface. The equipment was used by two classes of engineering students – one a general engineering introduction including civil engineering with 117 students, and one a specific civil engineering introduction with 234.

The two classes were operated separately, but with each expected to perform the same actions with the equipment – to set it to a given load and angle, observe the results, and draw conclusions based on determinate truss theory. Each class operated in groups of 3-4 students, completing their task as a piece of assessment. The laboratory was introduced to them as part of their regular tutorial over a 3-week period, though as it is remote students had the option to use it outside of the formal class times. The assessable portion of their work was due at the end of the second week.

The equipment is allocated to students through the Sahara Laboratories (Sahara Labs, n.d.) scheduling system, which includes a first-come-first-served queuing mechanism as well as pre-booking times. This allocation of the equipment is known as a ‘session’, which runs for a given length of time (15 minutes in this case). At the end of this allocated time, if no other students are waiting for the equipment or have a booking then a second 15-minute allocation is granted.

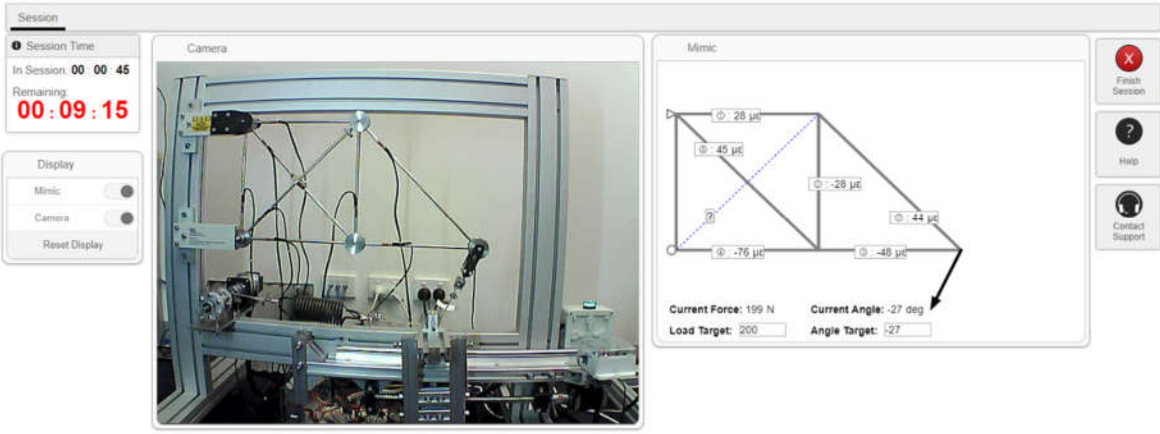

Figure 1: The truss laboratory UI

Dataset Selection & Collation

The system also provides queueing data indicating how long students had to wait to use the apparatus. Ad hoc observation of the system during the 3-week usage window revealed some large queue sizes lasting for multiple hours – however, in that time, the equipment was changing state very infrequently (i.e. no change to the load or angle). This raises questions about how students were making use of the apparatus and in particular suggests that students may been queuing to use the equipment for unnecessarily long periods. In order to improve this for future use of the laboratory, we reviewed previous work for possible solutions. Part of the existing research that has been conducted using Sahara Laboratories includes an examination of the queuing behaviours of students, and how technical changes to this system might impact the utilisation of the equipment (Lowe & Orou, 2012). The focus of that work was on the impact bookings had in preventing new sessions being allocated. Before implementing the most promising method from this paper (to allow for ‘shortened’ allocations where bookings clashed with immediate queue requests), we examined the available data from the Sahara system reports. This data showed a large queue size at certain windows as had already been observed, but booking behaviour did not show a similar impact to that reported in the prior research.

In addition to the queueing data the Sahara administrator interface can provide reports using an aggregation of data recorded in the backend SQL database that drives the scheduling and allocation system – but does not include all of this data. In addition, the system runs via an Apache web server that by default logs all URL requests made through it. By combining these two sources of data with our prior knowledge of the task students were given and their regular timetabled classes we have been able to collate enough information to gain insight into booking and queuing behaviour, and how it differed from the behaviour observed by Lowe & Orou (2012) with a different laboratory.

The data used includes:

- Session data – this includes a de-identified but consistent user ID, time of request, time of allocation (after any queue time), time of termination, reason for termination, and any associated pre-booking. We could calculate the number of students in the queue from a count of the session requests and the equipment state.
- URL requests – these are the specific requests made of the truss equipment itself, and include the time of the request, and the full URL. As the equipment is controlled via GET requests with data in the query string, this includes the type and value of commands sent. These commands change either the load or angle applied to the truss. By time comparison these can be associated with specific sessions and users.
- Class contexts – as described, we knew the tutorial times and assessment due dates, and we analysed the system's recorded data in light of these.

This dataset was assessed under the University of Sydney's human research guidelines as negligible risk, as indicated by the National Statement sections 5.1.22-5.1.23 and 2.1.7 (National Health and Medical Research Council (Australia) & Australian Research Council, 2018). Enabled by the use of a separate SAML-based institution authentication server not under the control or access of the researchers, no identifying data for any individuals was recorded in this dataset. After the conclusion of the semester prior to research commencing when the mapping table of SAML authorisation to user ID (an autoincrementing integer) expired and was deleted, no reidentification was possible. As a result, our use of this data is acceptable without explicit research consent from participants – indeed if a student did not opt-in or wanted their data removed, we would not be able to identify which records were involved. The data collected (and the laboratory software system itself) exists solely on infrastructure currently involved in other research that has required (and passed) stringent assessment for data privacy and loss protection.

Data

Queue size and booking behaviour

We first confirmed the difference seen in the standard booking reports; we examined cases where the student queue became high– each case coincided with scheduled class times (particularly where multiple classes aligned) and the weekend just before the assessment due date. Figure 2 is a snapshot of one of those scheduling periods with the number of students waiting (y-axis) over time (x-axis). Dark vertical lines show when the equipment was assigned to a new student and light vertical lines show cancellations.

The shaded areas indicate times when the equipment had a queue, but there was already a scheduled booking within the next15 minutes (the minimum allocation). It was these blocks of unallocated equipment that was the focus of the previous work in (Lowe & Orou, 2012). As in this figure, our analysis indicated that these had a very small in our case where there is demand in class times and near deadlines.

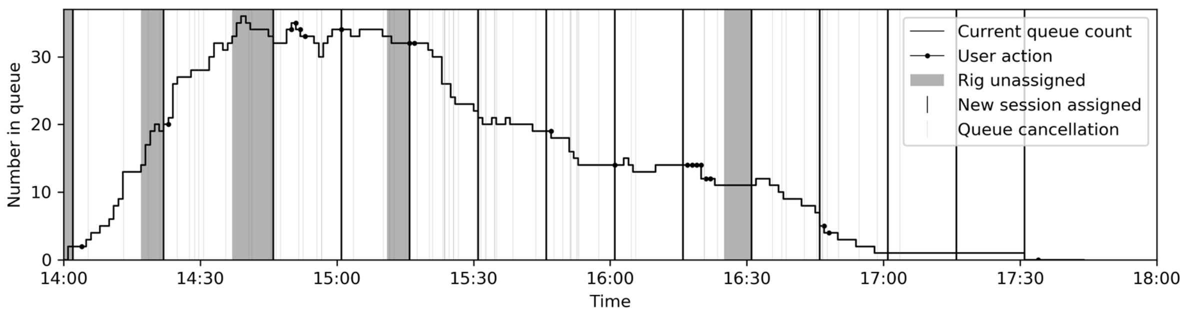

Figure 2: Queue size and booking behaviour over a 3-hour laboratory class (2pm-5pm)

Activity within user sessions

More interesting for our purposes were the large gaps in interactivity on the part of the users and equipment. The outputs from this data exploration showed clearly identifiable patterns of student behaviour when using the equipment. Figure 3 shows each recorded student session with the laboratory, sorted by time spent waiting in the queue (descending). *A higher resolution version of this and all other figures is available at https://ibb.co/album/gWSAka.* Each session starts on the left-hand axis at zero seconds and has a one second resolution, with vertical lines showing each minute. The markers on each session indicate points where the user gave a command (to change either the load or the angle), and then the reason the session was terminated, whether by the user's discretion or by running out of time when there was someone waiting in the queue or booking system, or if the user had used all their available time extensions (15 minutes base for a regular session, 30 minutes for a pre-booking, +15 minutes available once). If a user tries changing the angle while the system is under heavy load, the system will include a request to lower the load, then reapply it after the movement was completed – this appears in this record as as it represents a single human action

Some of the sessions show no actions at all – these appear likely to be users who left their computers or were working on a different task who missed their session entirely. The intended solution here is a simple prompt with a sound alert and timeout rather than automatic session assignment, to prevent both sessions assigned unnecessarily and dissatisfaction when students lose their place in the queue. We then studied the cases where users start a session, take an action, but then stop taking actions while remaining logged in. This chart only shows sessions where the student actively interacted with the equipment at least once – figure 4 will later explore the incidence of these 'no activity' sessions.

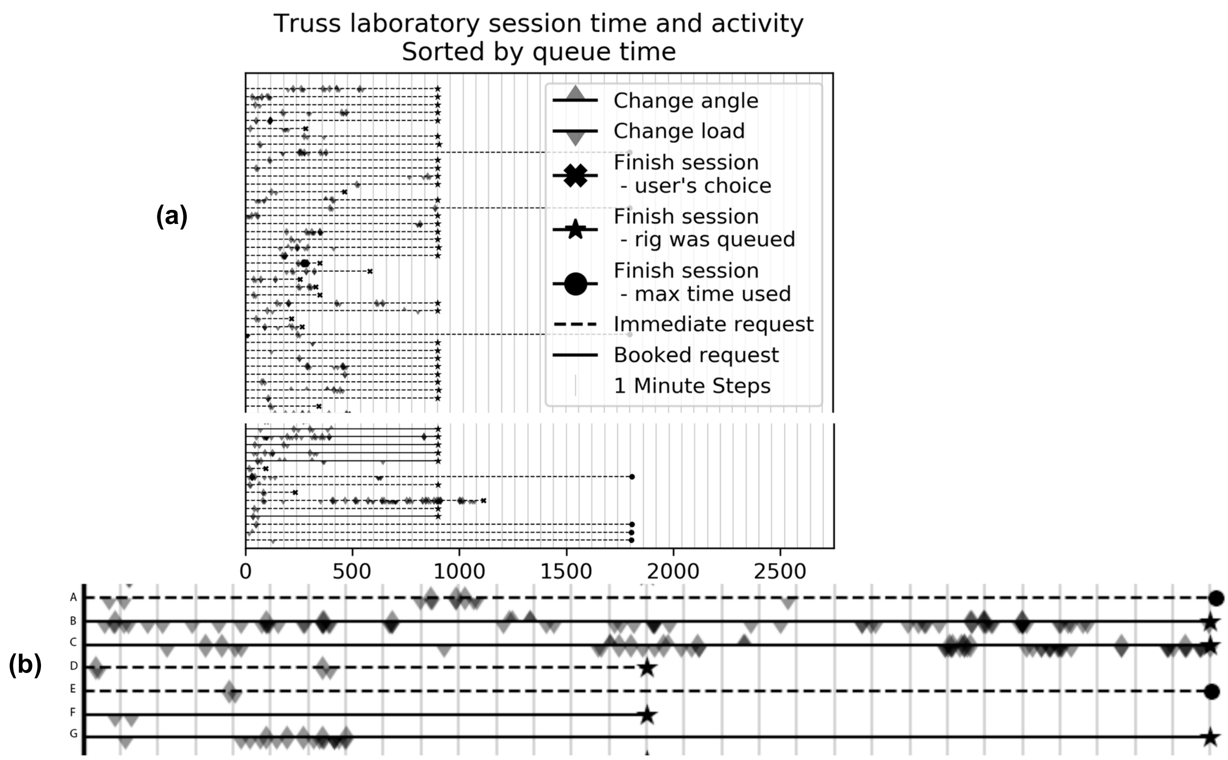

Figure 3: Activity within user sessions. (a) Extract from full tutorial class; (b) Zoomed-in details of 12 sessions illustrating diverse student interaction behaviours.

.

This figure matches the ad hoc observation that equipment often remained in the same state for long periods of time (see, for example, session G in Figure 3b) while it was assigned to a user. While the equipment takes at most 15 seconds to move from a loaded state at the minimum angle to a re-loaded state at the maximum angle, students for the most part log in to the system, enter a session, set the equipment to a single state (presumably the state they are given for their assessment) and then leave the equipment in that state while they observe the output – and most do not exit the session until the system terminates it.

There are, however, contrasting examples behaviour where students enter many different values throughout the length of their session (e.g. session B in the middle of Figure 3b). These examples mostly still show the enter-wait pattern that indicates they are allowing the system to arriving at a steady state before they enter their next command. There were some examples of students entering data in rapid succession (such as session C in Figure 3b) – without coming to a steady state the strain data would not be useful. It is possible these students are investigating the robustness and responsiveness of the equipment, which held up without problems. Our visualisation above makes it easy see these important types of uses of the equipment.

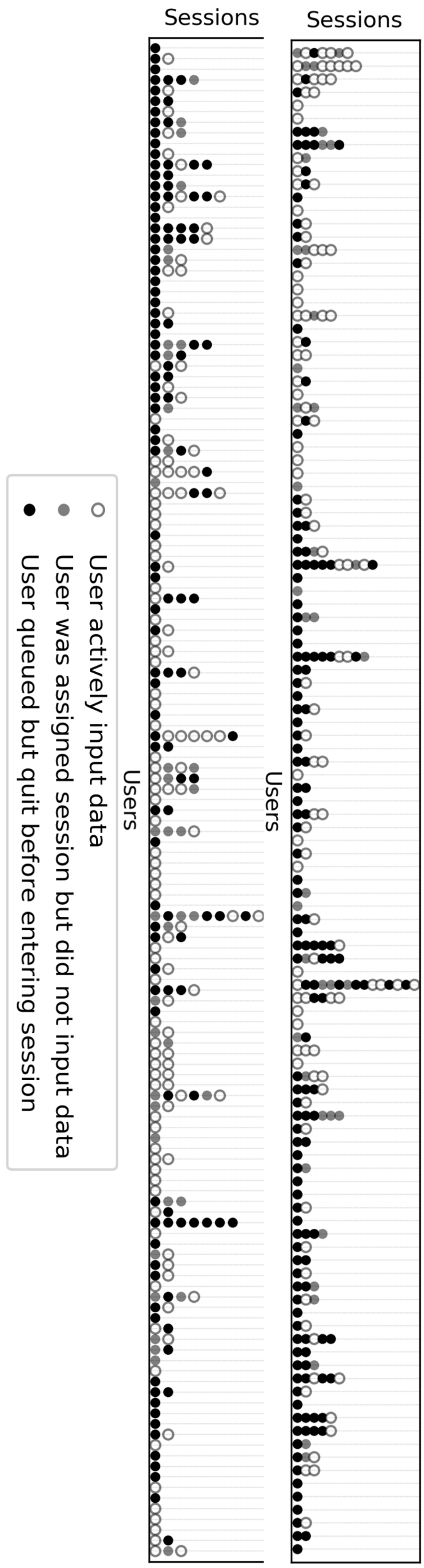

Figure 4: User session types

User session types

The preceding data shows there was both a significant queue delay during peak periods, as well as significant underutilisation of the equipment during these periods. However, feedback for the equipment was entirely positive from the lecturer and tutors – there were no complaints about heavy use or inability to get onto the equipment. This may be that in comparison to students' expectations of 'doing a laboratory' the necessary time and effort investment required to participate in this remote laboratory is low. In comparison, one traditional laboratory at the same institution requires students to book slots during several weeks at times outside any scheduled class - even with the peak queues observed on the truss student use mostly fit within already scheduled class time.

In order to investigate this underutilisation and inform improvements, we analysed the impact of the queue on the behaviour of students. Figure 4 displays the pattern of repeat session attempts by students over the period the laboratory was in use. These sessions extend upwards in chronological order, with each connected set of dots being a single user (split in half for ease of display). Of the 351 total students across the 2 classes 230 attempted a session and appear here.

The many black dots refer the very large number of queued students who quit before being allocated a session. The grey dots indicate a session where the student did not change the state of the equipment. This highlights the potential impact of equipment availability on the student behaviour:

- 54 users (only black dots) queued up, quit, but were never assigned a session with the equipment.
- 16 users (black then white dots) queued up, quit, and upon a subsequent session did not input data.
- 73 users queued up, quit, and later had a session where they input data.

The implication here is that 70 of the 230 students who tried to use the laboratory but were not given the opportunity within the level of time they were willing to spend to do so. Because the students were working in groups, however, it is possible that they worked together on a single computer within a session – hence the distinction of user rather than student and potentially explaining the gap between 351 total students and 230 observed users.

Preceding queue duration vs number of actions during session

The above figures each show trends that are easily distinguishable visually or by simple counts. They hint however at more sophisticated relationships between student behaviour and the system utilisation – and particularly the impact queuing has on the likelihood of a student to perform more actions, i.e. show more engagement with the system and interact with it in a more exploratory or inquiry-driven manner. Statistically significant relationships between queue length and number of events may be found – however this is simply extracting an entirely different correlation as illustrated in figure 5 below.

We were able to characterise session archetypes:

- Queue: None, Medium or High. None being a queue of zero, medium being 1-900 seconds (one 15-minute session) and high being more than 15 minutes.
- Use: Low, Medium or High. Low being 1-3 recorded interactions, which can cover one meaningful change in this learning context, medium being 4 to 20 recorded interactions, and then high.

This figure initially appeared to show entirely random behaviour, until the addition of session time categorisation. With the split between sessions that occurred during a scheduled class, sessions that occurred outside a scheduled class, and sessions on the weekend prior to the due date, the interaction of timing and behavioural categories become quite clear. Rather than being a function of queue time, the stronger predictor of the level of student activity in a session was whether they were accessing it during their regular tutorial or not. In addition, the impact of motivation on willingness to queue is apparent. Students were willing to queue for a session only up to a certain length of time during their tutorials, and this willingness increased with the approach of the due date of the assessment. Of those that waited longer than the initial time allotment of one session (1800 seconds) without cancelling, 22 were within the tutorial, 11 was outside the tutorial, but only 1 of those outside the tutorial was not close to the assessment due date. The willing time to wait increased to well over an hour in the most extreme cases near the due date.

The cluster of datapoints in this figure at the ‘low interactivity, zero queue length’ can be considered a baseline, with the interesting features appearing in the deviations from this common behaviour. With the exception of a single outlier in the centre of the figure where a session during a scheduled class had a significant queue length and significant interactivity, the highest queue lengths all occur close to the submission deadline, and the highest interactivity all occur outside scheduled classes.

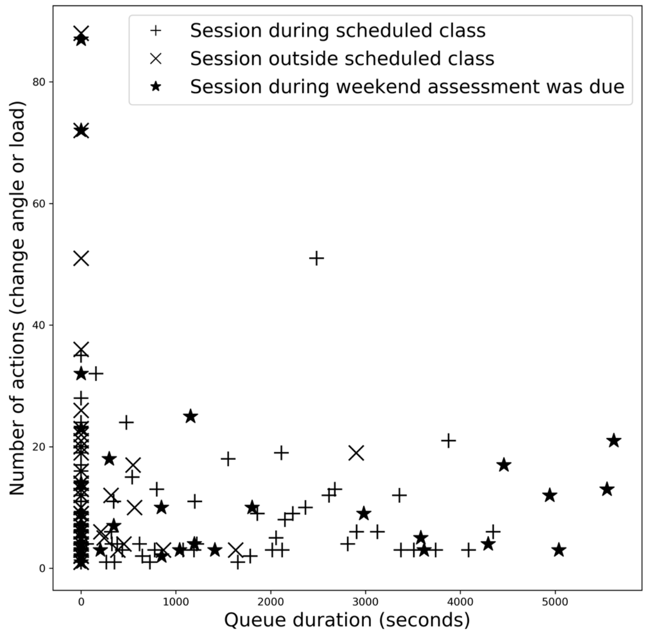

Figure 5: Preceding queue duration vs number of actions during session

Discussion

Laboratory Scheduling

The above data shows the limitations in the existing implementation of the laboratory scheduling, but also suggests opportunities to improve it. The steady-state nature of this laboratory means for the period the student is not actively changing states there is no need for the student to have exclusive access to the equipment. This characteristic in conjunction with the remote access of the laboratory means it may effectively be 'multiplexed' by allowing slices of time to be allocated to students' *actions* rather than assigning the equipment itself to students. Based on the observed use patterns this may bring down assigned session times from 15 minutes to ~1 in the case of single interaction sessions, for an improved utilisation of up to 1500%. Even taking the worst case as an average indicates a yield of 3-minute sessions and 500% of current utilisation.

This method of laboratory scaling already exists for remote electronic circuit laboratories where no visual media is used, such as described in (Gustavsson, 2002). As far as has been identified, this would be the first use of this method for a rich AV media remote laboratory such as the truss. Students will be brought into a virtual session without direct control of the truss and enter a value as they do now. If there is no queue, this value will be acted on immediately. If there is a queue, the student will still be assigned to a session and able to enter a value, but this value will be 'pending' until their place in the queue is reached. The strain data showing for the truss will represent the steady state after the student's last data entry was executed, allowing them to record their observations and discuss them with their group without tying up the equipment.

This particular modification is of course only possible due to the nature of the truss laboratory. Laboratories that required longer periods of interactivity, or that require interactivity where sequence or timing is important, or where there is no steady state to freeze, cannot be multiplexed in this way. There are, however, other methods this data collection has shown the potential of that may apply:

- showing the number of people waiting in the queue to motivate students to relinquish equipment they have finished using

- using more effective activity monitoring for the purposes of timeout behaviour
- requiring explicit ‘opt-in’ for longer session lengths when there is a queue would all have the potential to improve the utilisation of the truss rig.

Only through obtaining and analysing this data for a given laboratory can these possibilities be discovered and selected – and this is almost certainly the case for all laboratories to varying degrees.

Student Behaviour Modelling

This data collection has also yielded useful new insights about distinctive student behaviour. One of these involved a single setting then recording it. A very different behaviour involved exploring various settings. A third behaviour seemed to be testing the system. This is important in light of the established research showing the strong impact of student activity on their learning, both in the laboratory setting (Hofstein & Kind, 2012) and the wider field of education (Biggs, Kember, & Leung, 2001). Our approach harnesses available data without altering the learning context by making observations that change the learning context. It has enabled us to examine students’ use of laboratories to gain insights about student behaviour. This could be represented in a model of learners, similarly to work in educational data mining that has seen use in various e-learning activities (Romero & Ventura, 2010) – but does not yet appear to have been applied to laboratories.

Our approach opens the way to combine such models of learner behaviour with contextual data about what the educator expects student to do as well as analyses that include data about learning. This could be used to create adaptive tutorial material (Vandewaetere, Desmet, & Clarebout, 2011). the information may be provided to a tutor for intervention (Martinez Maldonado, Kay, Yacef, & Schwendimann, 2012). Appropriate forms of it could be displayed directly to the student as part of an open learner model (Bull & Kay, 2007). In this way the data available on student use of laboratories may be fed back in to improve their pedagogic effectiveness.

In the data in this particular case we can see a few broad if arbitrary categories of use: the one state change ‘set and record’ students; the repeated measurement ‘explorer’ students; and the rapid-fire changes students. These categories might be indicated either in real time or during assessment to educators or given as feedback to the students themselves. We note that additional qualitative data is needed to give meaning to these categories in the quantitative data (e.g. does rapid-fire input correlate with being unfamiliar with the required task? Does an explorer use pattern correlate with higher performance, and can such behaviour be better motivated in other students?).

Future Work

We believe this is report of this approach to data collection and analysis applied to remote laboratories. Our approach offers valuable opportunities for use in other remote laboratory systems. Indeed, it represents an exploration of ways to leverage the potential value of data that comes with any laboratory that has a computer-mediated interface. For example, moving the analysis of data from manual post-use generation to automated reports during the use of the equipment would enable feedback and intervention in real time. Our work also should inform the systematic design of new data collection that can help improve the operation and use of the laboratories and help teachers and learners gain insights from such data. This should include other types of laboratories, whether using existing locally operated computer mediated interfaces such as digital equipment via LabVIEW or adding IOT recording devices to regular equipment.

Our data analysis points to a promising approach to multiplexing the truss lab. It remains to explore the rich options in use of the video. Our work suggests promising options such as: to ‘freeze’ the video in place as it was after the effect of the student’s last input; to show the effect of other students' use in real-time; or offer both. Pedagogic arguments could be made either way (a sense of ownership for the first, a sense of being part of an active community of learning in the second, etc), a sense of control in the third. Future work will be to investigate which of these choices are helpful.

Conclusion

Computer mediated interfaces to laboratories allow the collection of student interaction data that was previously unobtainable or could only be collected by observations that altered the learning experience. This paper has shown that by leveraging this "passive observation" data we can learn more about students' interaction with laboratories in order to inform changes to pedagogic and system design. The specific case of the truss has yielded insight into its operation. It has also provided rich insights to motivate and inform improvements to the scheduling mechanism and interface. By obtaining this data without needing to query students for it directly or to use obtrusive or invasive observation techniques, the ecological and internal validity of the data is drastically improved ('defeating Hawthorne') – and its computer-driven nature means real-time systems may be developed so that they can make use this data as an intervention in itself. The data used is likely to be readily available in other remote laboratory systems or indeed any laboratory using a computer-mediated interface, and thus these methods may be applied widely.

References

Biggs, J., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor Study Process Questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, *71*(1), 133–149. https://doi.org/10.1348/000709901158433

Bull, S., & Kay, J. (2007). Student models that invite the learner in: The smili open learner modelling framework. *International Journal of Artificial Intelligence in Education*, *17*(2), 89–120.

Gustavsson, I. (2002). Remote laboratory experiments in electrical engineering education. *Proceedings of the Fourth IEEE International Caracas Conference on Devices, Circuits and Systems (Cat. No.02TH8611)*, I025-1-I025-5. IEEE.

Hofstein, A., & Kind, P. M. (2012). Learning In and From Science Laboratories. In *Second International Handbook of Science Education* (pp. 189–207). https://doi.org/10.1007/978-1-4020-9041-7_15

Lowe, D., & Orou, N. (2012). Interdependence of booking and queuing in remote laboratory scheduling. *2012 9th International Conference on Remote Engineering and Virtual Instrumentation (REV)*, 1–6. https://doi.org/10.1109/REV.2012.6293100

Martinez Maldonado, R., Kay, J., Yacef, K., & Schwendimann, B. (2012). An Interactive Teacher's Dashboard for Monitoring Groups in a Multi-tabletop Learning Environment. In *International Conference on Intelligent Tutoring Systems 2012 (ITS 2012)* (pp. 482–492). https://doi.org/10.1007/978-3-642-30950-2_62

National Health and Medical Research Council (Australia), & Australian Research Council. (2018). *National Statement on Ethical Conduct in Human Research (2007) - Updated 2018*.

Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the Worker*. Oxford, England: Harvard Univ. Press.

Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *40*(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Sahara Labs. (n.d.). Sahara Labs GitHub Repository. Retrieved June 21, 2018, from https://github.com/sahara-labs

Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, *27*(1), 118–130. https://doi.org/10.1016/j.chb.2010.07.038

Copyright statement