COLEG GWYDDORAU'R AMGYLCHEDD A PHEIRIANNEG
COLLEGE OF ENVIRONMENTAL SCIENCES AND ENGINEERING

YSGOL CYFRIFIADUREG A PHEIRIANNEG ELECTRONIG
SCHOOL OF COMPUTER SCIENCE AND ELECTRONIC ENGINEERING

PRIFYSGOL BANGOR UNIVERSITY

JISC/Bangor University Learning Analytics Project Summary & Case Study

Dave Perkins, Cameron C. Gray, Panagiotis D. Ritsos, and Ludmila I. Kuncheva

The School of Computer Science and Electronic Engineering at Bangor University propose launching a new project within our established Learning Analytics Initiative. The Initiative is currently hosted by the School with guidance from the University's Centre for Enhancement of Learning and Teaching (CELT).

Abstract

Insights into activities we undertake as educators and students have the potential to enhance learning and reduce unintentional consequences for all. Educators have for a long time used data to monitor students and grade them. More recently additional yet still traditional metrics have been added to the available tools in every day education. The latest generation of information are derived metrics with additional intelligence. This project has developed a *Work Pressure* metric than can be used by both educator and learner. The focus is on the assessments for a given programme and Work Pressure that this generates. Additionally, included is behavioural characteristics, these have the potential to have significant impact upon the individual student journey.

Project Description

While Learning Analytics are frequently used to keep a watchful eye on student performance and engagement; one area that has yet to be monitored is students' workload. This workload has been noted as a significant factor contributing to students' deteriorating mental health[1]. This fits with one of JISC's priorities, looking at analytics to assist with Student Welfare and Mental Health. Previous studies[2] have already shown that there is a correlation between workload, stress and performance. However, time management is not the only factor[3] in determining performance. Workload has also factored prominently in studies of student satisfaction [4], [5].

The concept of assessment and course design is often completed using large spreadsheets. Through no deliberate fault of the designers, often only the deadline dates are considered. Individual instructors are then left to decide suitable time frames for the assessments they are setting. Quality procedures then check for clusters or overlaps of deadline dates. As a result, educators do not sufficiently consider the overall workload for students (e.g. the amount of time/effort required or complexity of the activities). Due to the UK HE External Examiner system, we do not believe this to be an institutional phenomenon.

In order to narrow the gap between tutors' expectations and students' impressions, module leaders would benefit from an understanding of the existing workload when designing and timing assessments. This goal will require a comprehensive and easy-to-follow tool. As not every student is the same, this tool must be able to model various student responses to another assessment released. This form of analytics would allow curriculum designers to run 'what if?' scenarios before setting more assessments that could potentially be detrimental to mental health or morale.

PRIFYSGOL BANGOR
BANGOR, GWYNEDD,
LL57 2DG, DU

+44 (01248) 351151
WWW.BANGOR.AC.UK

BANGOR UNIVERSITY
BANGOR, GWYNEDD,
LL57 2DG, UK

+44 (01248) 351151

Project Outputs

The project developed a staged set of outcomes, starting with a metric to represent the Work Pressure, this was then used in the development of a visualisation.

Output 1: Descriptive Metric of Student Work Pressure

The descriptive Student Pressure Metric (P) is a synthetic metric, defined at any point in time t, with an open set of assessments $A(t)$. Each item (denoted as item i) in set A has a start time $S(A(i))$, an end time $E(A(i))$ and an expected effort $X(A(i))$. The effort expended by the student so far on the assessment is denoted as $F(t, A(i))$ and its calculation will depend on the behavioural model in use. We also define W as the amount of expected working time. It is important that E, S, t, X, F and W are all expressed in the same measure/units of time, whether that is at day or hour resolution is unimportant for the mathematical definition. Equation (1) defines P mathematically.

$$P(t) = |A(t)| + \sum_{i=1}^{|A(t)|} \frac{X(A(i)) - F(t, A(i))}{(E(A(i)) - t) \cdot W} \tag{1}$$

Essentially this is an amalgam of the number of deadlines that students have to contribute toward and a ratio of the amount of time left to the amount of expected effort remaining. This takes into account both static pressure, measured as the amount of assessments/deadlines, and dynamic pressure of effort still to be completed that is observed in the relevant literature [6], [7].

The prototype tool/system has defined a single behavioural model, named 'Consistent'. This type of student diligently starts work on an assessment when it is set and works until the deadline in a linear fashion. For a single assessment this can be defined mathematically as in Equation (2) using the same symbols already defined.

$$F(t, A(i)) = (t - S(A(i))) \cdot \frac{X(A(i))}{E(A(i)) - S(A(i))} \tag{2}$$

Equation (2) is effectively a cumulative linear scaling function of expected effort over the duration of the assessment. The Project Team acknowledges that while this behavioural model is the easiest to define mathematically, and therefore implement, it is probabilistically the least likely to occur in a student body. Future work would complete four additional behavioural models named 'Keen', 'Stuck-in-the-middle', 'Stuck at Start', and 'Late Starter'. Idealised versions of these curves can be seen in Figure 1.

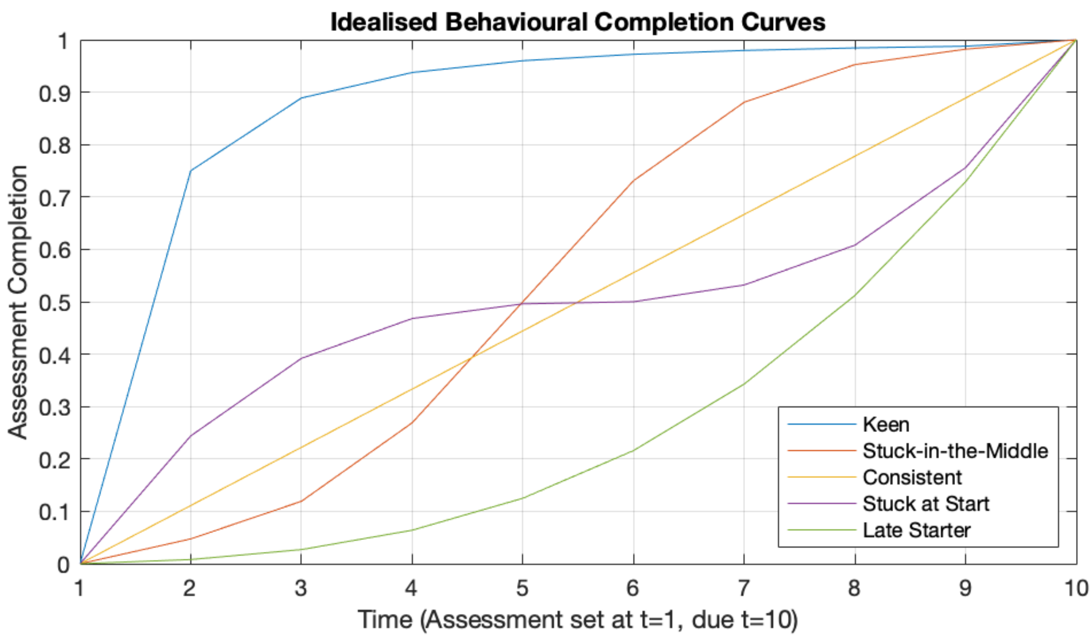


Figure 1: Idealised assessment completion curves for all behavioural models. The X-Axis shows time (units largely irrelevant), and the Y-Axis shows the completion as a decimal percentage.

As the dynamic basis of this metric is based on time, a normalisation exercise is needed across programmes, modules, student work ethics, and even institutions. Many UK institutions use the Module Credit system, where each year of study is divided into 120 credits [8]. This same system standardises a notional value of 10 hours of study per credit. Effectively this sets 1200 hours per year as the amount of time an average student is expected to spend on their studies. There is some discussion [9] as to how universally applicable this standard is. However; in the absence of another accepted and widely used measure, this project has adopted this as the basis of 'time' for student work. The one remaining variable to convert the notation hours into practical time is how long an institution's semesters are. For example; Bangor uses two twelve week teaching semesters, leading to $1200/24 = 50$ hours per week study time. The University of Hertfordshire has 15-week semesters [10], leading to $1200/30 = 40$ hours per week study time.

The notation learning time includes all aspects of a student's course; compulsory scheduled events such as lectures and practicals as well as private study. These compulsory events normally have set learning objectives of their own rather than directly contributing time to assessments. Therefore, this time is deducted from a student's available assessment time. It is this assessment time that is used in the metric calculations. Anecdotally, lecturers observe a lower attendance at scheduled events around assessment deadlines. This finding is supported by at least one study [11], however it is not universal nor consistent enough an effect to be built into the time calculations.

Within Bangor University, there is a confounding factor with using expected effort/time. The Quality Assurance and Validation Unit have published an Assessment Framework designed to harmonise the experience of students assessment across the institution. This document* sets out a guideline of 2-2.5 hours' effort towards assessment per module credit. The prototype system includes both measures of expected effort as the authors believe that educators design their assessments proportionally to the weighting of marks, rather than on this standardised scale. i.e. The assessment is worth 40% of the course therefore should take approximately 40% of the assessment time available. In the cases examined, and for the most part, there is a linear scaling between the two measures. This results in pressure points being hidden behind a smaller range, potentially misleading programme designers.

The prototype tool includes both measures of time for comparison purposes, which can be useful when providing evidence of parity during course validation efforts. Due to the lower amount of effort indicated by the Framework and the division used within the Pressure Metric, using the Framework measures results in a numerically higher Pressure Metric value. This is expected, and the two time measures are not intended to be directly/numerically compared.

Outputs 2 & 3: Prototype Data Manipulation and Visualisation Tool

Drawing on the authors' own experience of programme design, coincidentally as part of the School's most recent course redesign, the visualisation of the pressure students are under has become a series of co-ordinated views [12] from different perspectives. These follow the, now famous, Shneiderman mantra [13]; overview first, then detail on demand.

The overview takes the form of a time-series bar chart showing the number of active assessments per day and a set of summary statistics. This overview is pictured in Figure 2. Each bar is coloured to give an indication of potential issue areas. These colours are linearly gradated from green to red, the exact colours are dynamically set according to the loaded data set. This scheme has obvious advantages, the colour-coding instantly draws the viewer to areas with significant work and shows duration of the static pressure on the student. However, with some designs of course where numbers of active assessments are tightly controlled this view loses its utility. This is because there is no thresholding or added intelligence for what is considered an acceptable number of assessments. This is an element that can be added very simply, subject to analysis of a wide range of courses and potential collaboration with educationalists, psychologists, and course designers. It is unknown at this stage if this would need to be a threshold set at the department or institution level.

The textual statistics give a quantitative view of the same overview data. These include start and end dates for the assessment year, which may differ significantly from the academic year in some places. It also shows raw counts of assessment and the breakdown of the notional learning hours between the compulsory scheduled teaching events and the remainder to be used for assessment. These statistics should be comparable across programmes, however it is unclear how useful this would be outside of a single department.

The second view allows course designers to judge individual assessments in the wider context of the assessment year. This uses a Swimlane chart showing horizontal bars for each assessment. The view collapses the number of rows to a minimum possible, which matches the maximum number of assessments at any one time. These bars are coloured by module, the actual colour value has no special significance. An example view is shown in Figure 3.

The duration view is comprised of two elements a full-year element (bottom) and a detail element (top). The full-year element uses a technique called Brushing [14], allowing the user to select a region (a period of time in this case) to focus on. The shaded blue/purple shows the current focus period. The default is a two-week window

*https://www.bangor.ac.uk/quality/course/documents/BUAssessmentFrameworkOct2018.pdf

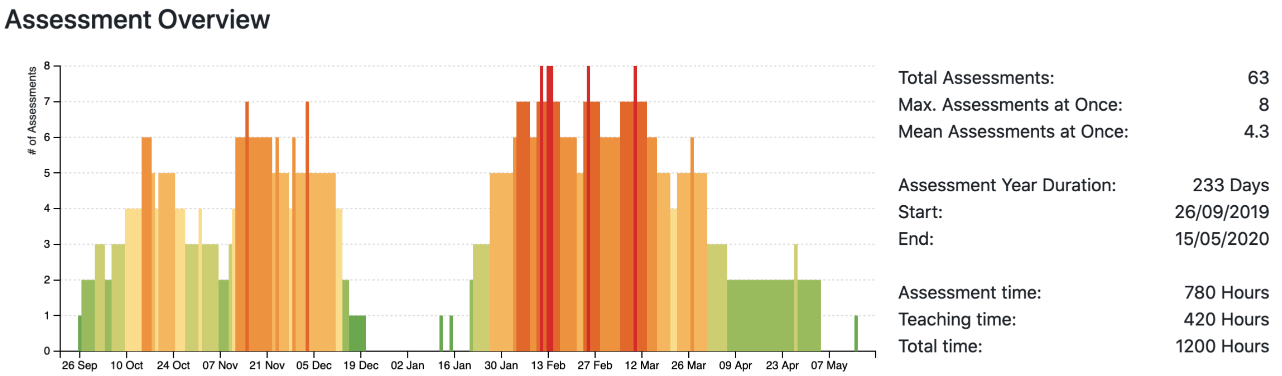


Figure 2: The Assessment Overview view from the prototype system. Comprised of linearly-gradated bars showing number of assessments alongside summary statistics.

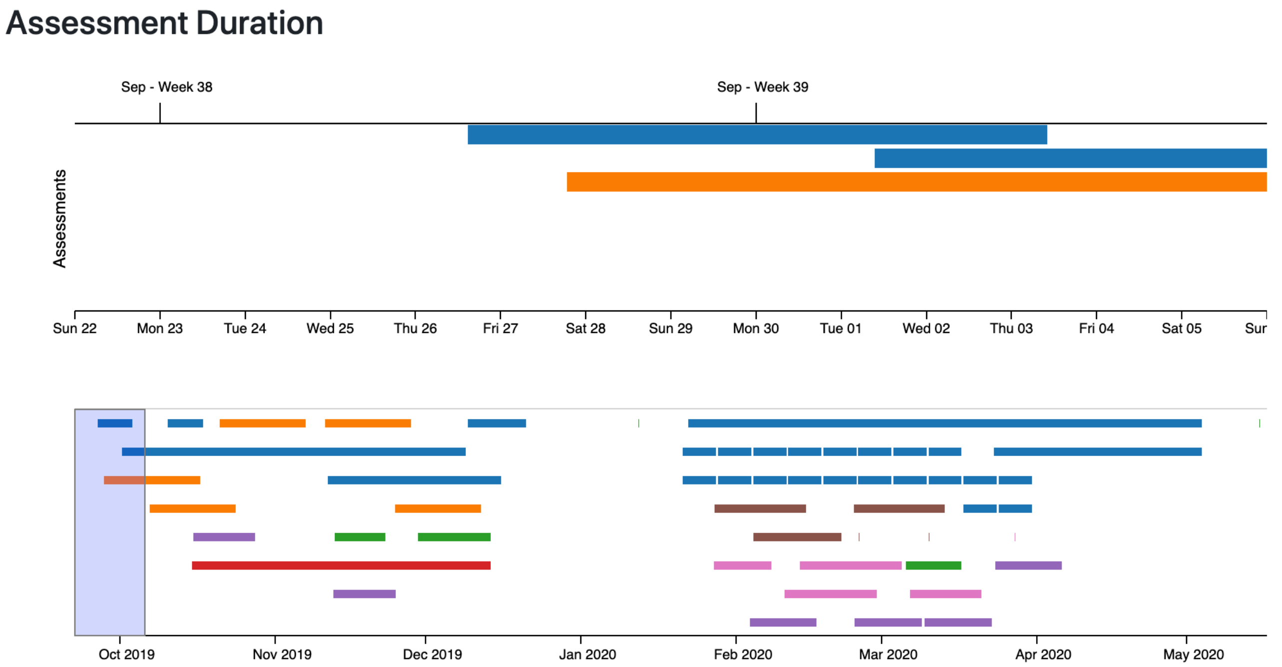


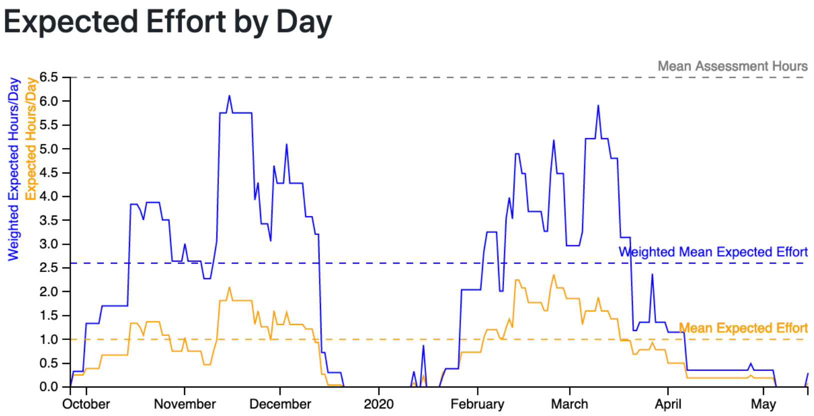
Figure 3: The Assessment Duration view using a Swimlane chart. The bottom element shows the full year overview which is fixed. The top element shows the full resolution period selected by the user.

but this can be resized. Hovering the pointer over elements in the detail element provides the details of which module and assessment is shown.

The remaining views are affected by the selection of a behaviour model, and depict the outcome of the design in terms of absolute effort hours required and the Pressure Metric. These are shown in Figure 4 and Figure 5. For the absolute effort hours, the view includes mean lines to show the mean across the non-zero year. This is to remove holidays such as Christmas and Exam Weeks from distorting the calculation. It also includes the level of effort per day expected from the notional credit hours calculation. This is useful for designers to see if at any point they would be expecting students to commit more time than they had available.

These views, along with the Assessment Overview view, also have interaction where clicking on a part of these graphs will refocus the Assessment Duration view to the appropriate date. This allows users to quickly identify potential problem assessments or areas by date.

At present there is no UI to conduct the 'What If?' experiments, however the system is based on two plain-text files containing the description of programmes and assessments in JavaScript Object Notation (JSON) format. These files can easily be manipulated to conduct these forms of experiments, albeit somewhat manually. This common data model would be the basis for transforming data from existing LMSs, course and campus management systems. In order to correctly handle the what-if scenario this data would need to be duplicated for each experiment into some form of scratch (or disposable) database or files. The handling and sharing of these scratch forms would need careful management in a multi-user and even more so in a multi-tenant system such as JISC's Learning Analytics platform.



Mean Teaching Hours/Day	3.5
Mean Assessment Hours/Day	6.5
Mean Expected Effort Hours/Day	1.0
Mean Weighted Expected Effort Hours/Day	2.6
Max. Expected Hours/Day	2.4
Max. Weighted Expected Hours/Day	6.1

Figure 4: The Expected Effort view including quantitative statistics for both time/effort measures with links to the notional credit hours. Means are calculated from non-zero values.

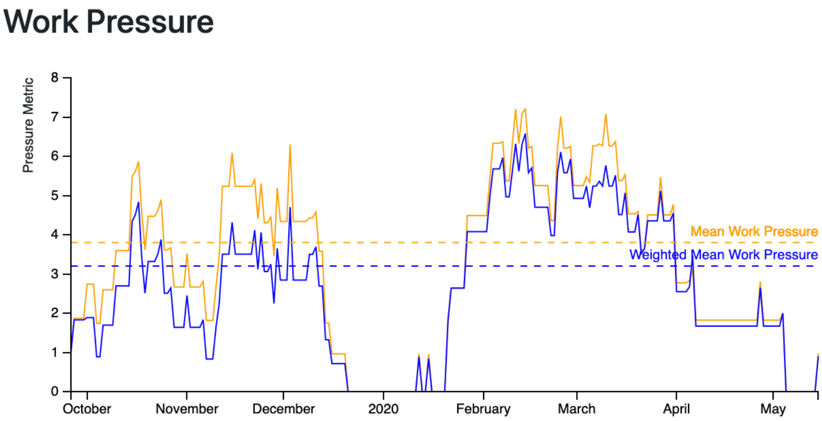


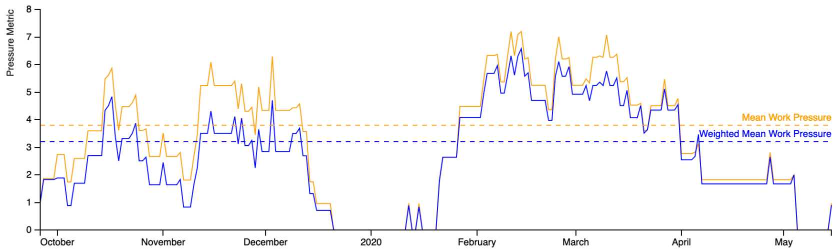
Figure 5: The Work Pressure view showing the resulting Pressure Metric value over time. This view shows the calculation with both time/effort measures and a mean level calculated from non-zero values.

Case Study 1: Comparison between Courses

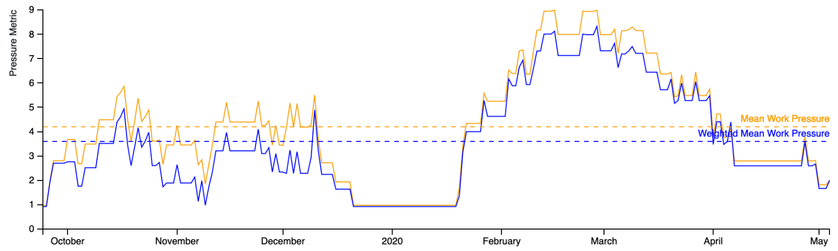
Bangor's School of Computer Science and Electronic Engineering offer two main CS courses; the flagship Computer Science programme and the more applied Computer Information Systems programme. These programmes offer two different approaches: CS is designed to be theory led asking students to understand how computers work; CIS how to achieve end results using computing resources. As such there are two different sets of expectations for students on the respective courses. One deals with abstract theory and levels of minutiae requiring the student to understand and apply that knowledge for themselves. The other presents material in sets of steps, and relies on repetition with minor alterations to develop understanding in the students. In short they could be summarised as 'hard material, less time-consuming work' and 'easier material, applied multiple times'.

The Prototype Tool illustrates this well, seen in Figure 6. The CS student pressure is more volatile and 'peaky' as assessments start and finish. This characterises lots of smaller assessments with significant individual learning in between. The CIS student pressure however is more sustained, this indicates a steady flow of assessment over the year, including long-running assessments such as projects and portfolios. There are also three spikes set aside from the rest of the curve for CS. Two occur in mid-January, and the other late May. These are traditional, formal exams. As a result of the different aims of the programmes CS still includes the theoretical exams, where CIS features more continuous assessment.

However, the inclusion of the Mean lines (for both time models) reveals that both programmes apply a roughly equivalent amount of pressure on the students. This was a purposeful decision on the part of module designers, to combat the idea that one degree was 'easier' than the other. This design was put in place prior to the development of the prototype, so it is seen as a confirmation of the design.



(a) Computer Science, Year 1.



(b) Computer Information Systems, Year 1.

Figure 6: Comparison of Work Pressure between Bangor's two main CS programmes. Note that the Mean Pressures are similar, and that CIS students have a more sustained load across the year.

Case Study 2: Highlighting Less Experienced Educators

All departments have newer or less experienced educators alongside those with more experience. This division is also possible between Teaching & Scholarship lecturers and their Teaching & Research colleagues. The divide is unavoidable, but it may also be limited to educational design rather than delivery or tenure in post.

Assessment design is acknowledged to be an intricate and challenging task [15]. There are no universal methods applicable to every subject and cohort, requiring educators to constantly adapt. This requirement for adaptation is where Learning Analytics comes into its own. The output of the analytics process is able to give the educator near-real-time feedback on whether their design and strategy is working. However, this only happens once students are engaging with the course.

Using our Assessment Duration view from the prototype, managers and course leads can see a clear difference in assessment strategies. Figure 7 shows one such example. There are two distinct sets of assessments. Those in blue and brown, two modules led by less-experienced educators; and the remaining assessments marked in green, pink and red.

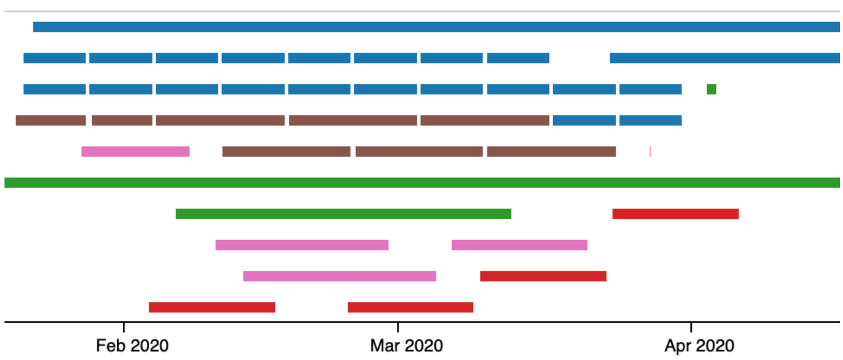


Figure 7: The assessment duration view of semester two for the Computer Information Systems Year 1 cohort. There are two distinct strategies split into the blue and brown group, and the others.

The module organiser for those marked in blue, has a long running group assessment and decided that it was necessary to monitor progress on a weekly basis. To that end, they instituted an individual and group diary exercise worth relatively little in terms of marks (1% per submission). As the expected effort on these pieces of work are commensurately low, they exert little dynamic pressure on the student. However, the static pressure of another deadline is significant.

The second example, again a new educator, is the module shown in brown. In this case the module leader introduced a new topic each week which included an assessment. Students were given that week's laboratory time and an additional week before needing to submit their work. This creates an overlap of assessments, once again adding to the static pressure on their students.

The intentions and rationales for these designs are understandable and in isolation entirely reasonable. However, this highlights one common failing among newer educators: they tend to operate in isolation, not taking the wider programme into consideration. While the prototype will not prevent these situations from occurring, it does provide situational awareness for both programme leads and individual lecturers. Both roles can see where peak/pinch points occur, and provide empirical evidence for adjustments being neccessary.

Project Review

This study was originally intended to be conducted over a six month period, this was to allow for the university to make the data available and development of the tool and models.We set out a number of milestones in the proposal in order to complete the project, these are detailed below:

Milestone 1 Gained access to the raw data - end of month 1.
Milestone 2 Scrubbed and Examined the data - end of month 2.
Milestone 3 Developed a working metric from the data and hold a focus group with HE Lecturers to devised an initial workload model - end of month 3.
Milestone 4 Have implemented the visualisation tool and start to perform the qualitative evaluation process - end of month 4.
Milestone 5 Have a final tool in place for second focus group - end of month 5.
Milestone 6 Have written an article/case study for publication - end of project.

These milestones we're intended as guidelines but also identified the critical path to success of the project. *Milestone 1* would seem to just be a technicality to move the project forward, it has fundamentally not been met, there were numerous delays in gaining access to the data, then assumptions made that led to the data being manually extracted from the systems. Once completed the project progressed swiftly. Aside from the data access provision the project went well, there is however considerable scope for future development and a hope deployment to help colleagues in planning and curriculum design.

Future Work

Future Work would surround implementing, subject to appropriate collaboration, the remaining student completion models. Additionally it should be possible to weight the likelihood of these models based on observations of representative cohorts. This is likely to give an indication as to which side educational designers should err toward. Based on anecdotal evidence it is likely to give rise to an impossible choice. Real students would tend toward the extremes of these behavioural scale rather than cluster in the centre. This has implications for all curriculum design, should educators design more for High Flyers ('Keen' students in our terminology) or those that leave everything to the last possible moment? Which group is it better to moderate the pressure on? Which group would benefit most from this sort of accommodation?

From that point there is two possible directions to travel toward. The first is integrating more emotional and soft aspects of student life into the model. This will require input from psychologists and social scientists to be able to quantify these stresses and to identify what part of the academic year they fall into. The second is converting this model into a self-service tool for student themselves. Allowing students to not only select their behaviour type but add in part time employment, social and extra-curricular commitments in order to produce as representative view of their likely load/pressure. In addition there may be worth in including weightings that students can select by assessment type to alter the static pressure 'value'. This will allow the personalised model to adapt to the fact that some students excel in practical work but find tests and exams stressful, and vice-versa.

References

[1] J. Hunt and D. Eisenberg, "Mental health problems and help-seeking behavior among college students," *Journal of adolescent health*, vol. 46, no. 1, pp. 3–10, 2010.

[2] S. E. Ross, B. C. Niebling, and T. M. Heckert, "Sources of stress among college students," *Social psychology*, vol. 61, no. 5, pp. 841–846, 1999.

[3] E. A. Plant, K. A. Ericsson, L. Hill, and K. Asberg, "Why study time does not predict grade point average across college students: Implications of deliberate practice for academic performance," *Contemporary Educational Psychology*, vol. 30, no. 1, pp. 96–116, 2005.

[4] S. Tresman, "Towards a strategy for improved student retention in programmes of open, distance education: A case study from the open university uk," *The International Review of Research in Open and Distributed Learning*, vol. 3, no. 1, 2002.

[5] J. Douglas, A. Douglas, and B. Barnes, "Measuring student satisfaction at a uk university," *Quality assurance in education*, vol. 14, no. 3, pp. 251–267, 2006.

[6] D. S. Ackerman and B. L. Gross, "Is time pressure all bad? measuring the relationship between free time availability and student performance and perceptions," *Marketing Education Review*, vol. 13, no. 2, pp. 21–32, 2003.

[7] N. Paden and R. Stell, "Reducing procrastination through assignment and course design," *Marketing Education Review*, vol. 7, no. 2, pp. 17–25, 1997.

[8] "Higher education credit framework for england:guidance on academic credit arrangementsin higher education in england," The Quality Assurance Agency for Higher Education, Gloucster, United Kingdom, 2008. [Online]. Available: `https://www.qaa.ac.uk/docs/qaa/quality-code/academic-credit-framework.pdf?sfvrsn=940bf781_12`.

[9] S. Lester, "The uk qualifications and credit framework: A critique," *Journal of vocational education and training*, vol. 63, no. 2, pp. 205–216, 2011.

[10] University of Hertfordshire. (2016). "Study-load: Understanding and managing your studies," [Online]. Available: `https://catoolkit.herts.ac.uk/toolkit/study-load-understanding-and-managing-your-studies/` (visited on 05/12/2020).

[11] M. Hitchens and R. Lister, "A focus group study of student attitudes to lectures," in *Proceedings of the Eleventh Australasian Conference on Computing Education-Volume 95*, Australian Computer Society, Inc., 2009, pp. 93–100.

[12] J. C. Roberts, "State of the art: Coordinated & multiple views in exploratory visualization," in *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)*, IEEE, 2007, pp. 61–71.

[13] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Proceedings 1996 IEEE symposium on visual languages*, IEEE, 1996, pp. 336–343.

[14] D. A. Keim, "Information visualization and visual data mining," *IEEE transactions on Visualization and Computer Graphics*, vol. 8, no. 1, pp. 1–8, 2002.

[15] M. van den Heuvel-Panhuizen and J. Becker, "Towards a didactic model for assessment design in mathematics education," in *Second international handbook of mathematics education*, Springer, 2003, pp. 689–716.