International Journal of Mathematical Education in Science and Technology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tmes20

The effect of incentivising pre-class reading on engagement and student performance

Amanda J. Shaker, Christopher Brignell & Mathew Pugh

To cite this article: Amanda J. Shaker, Christopher Brignell & Mathew Pugh (2023): The effect of incentivising pre-class reading on engagement and student performance, International Journal of Mathematical Education in Science and Technology, DOI: 10.1080/0020739X.2023.2214562

To link to this article: https://doi.org/10.1080/0020739X.2023.2214562

Published online: 30 May 2023.

INTERNATIONAL JOURNAL OF MATHEMATICAL EDUCATION IN SCIENCE AND TECHNOLOGY
https://doi.org/10.1080/0020739X.2023.2214562

The effect of incentivising pre-class reading on engagement and student performance

Amanda J. Shaker [a], Christopher Brignell [b] and Mathew Pugh [c]

[a]Department of Mathematical and Physical Sciences, La Trobe University, Melbourne, Australia; [b]School of Mathematical Sciences, University of Nottingham, Nottingham, UK; [c]School of Mathematics, Cardiff University, Cardiff, UK

ABSTRACT

A commonly-held belief is that many university students are 'assessment-driven'; that is, students engage more with activities that are assessed compared with those that are not. 'Incentivised engagement' includes the practice of providing incentives (such as marks or otherwise) for students to engage in particular activities. Perusall is a social annotation platform designed to increase student engagement with pre-class reading. In this study, students in a second-year undergraduate Statistics for Engineering module were invited to carry out the weekly reading activity via the Perusall platform. During the first half of the semester, students' engagement via Perusall did not count towards student marks, whereas during the second half of the semester, it did. We present the estimated effect of incentivised engagement, via use of the Perusall platform, on student engagement and performance. Our results show that incentivisation increased the engagement of students who were previously less engaged with the pre-class reading. For those students who were previously less engaged and had lower performance in summative assessments, there was also an increase in student performance following incentivisation. However, this effect was not observed for students who were already performing well before incentivisation.

ARTICLE HISTORY
Received 27 June 2022

KEYWORDS
Incentivisation; engagement; perusall; student performance; assessment

1. Introduction

A commonly-held belief is that many university students are 'assessment-driven' (e.g. Holmes, 2018); that is, students engage more with activities that are assessed compared with those that are not. 'Incentivised engagement' includes the practice of providing incentives (such as marks or otherwise) for students to engage in particular activities. However, there is very little evidence about the impact of incentivisation, and the results that exist are mixed.

Perusall (https://perusall.com/) is a social annotation platform where students make annotations on a shared document and can view and reply to annotations posted by their peers and teachers. It is designed to increase student engagement with pre-class reading. It

CONTACT Amanda J. Shaker a.shaker@latrobe.edu.au Department of Mathematical and Physical Sciences, La Trobe University, Melbourne, Australia

measures student interaction with the text by, amongst other things, monitoring how long a student spends actively reading and assessing the number and quality of annotations a student makes. Instructors can then choose whether Perusall's scores contribute to a subject's assessment. The validity of the scoring system is discussed further in Section 2.2.

The aim of this study is to investigate the effect of incentivising students taking a second-year probability and statistics for engineering module, to engage in pre-class reading using the Perusall platform. In particular, we ask, 'does incentivisation lead to a higher level of engagement with the task?' and, if so, 'does incentivising the task lead to higher student performance?' As discussed in the following section, the incentivisation literature suggests that the answer to the first question is 'yes' but that the answer to the second question is more nuanced. Therefore, this study also aims to understand the types of students who may benefit more from incentivisation. The Perusall literature (discussed in Section 2.2) implies its use will enhance performance through students gaining deeper conceptual understanding. However, no study has investigated the role incentivisation plays in Perusall's apparent success, or its use in teaching statistics. Therefore, this study also aims to identify whether incentivisation is an essential component in deploying Perusall, particularly in the context of a statistics module.

2. Literature review

2.1. Incentivised engagement research

In recent years student engagement has been paid increasing attention, including several literature reviews, (e.g. Trowler, 2010; Wimpenny & Savin-Baden, 2013) and special issues of journals (Macfarlane & Tomlinson, 2017). Many studies have found correlation between higher levels of student engagement and academic grades (e.g. Indiana University Center for Postsecondary Research, 2002).

An increasingly common approach to encourage higher levels of student engagement is to incentivise this through external rewards. Anderson (2016) tried to design a course structure where the incentives (rewards) for engagement with formative assessment were access to lecture notes before the lecture rather than having to wait until after the lecture. The aim was to design a system which would not disadvantage academically those who did not engage, but to reward and provide a sense of competence and autonomy to those who did. Students who engaged with these activities reported higher engagement, confidence, preparedness and subject enjoyment than those who did not. Another approach is simply to monitor engagement and contact students whose engagement is not satisfactory. Burke et al. (2013) found that this had a positive impact on students' engagement during the first semester, but had little impact in the longer term.

However, the most common reward used is the award of marks which contribute towards the course grade. This is what will be meant by incentivisation in this article.

Not surprisingly, incentivising engagement leads to higher levels of engagement, at least in terms of the activity being incentivised. Beard (2017) introduced regular summative e-assessments and found that student engagement was high in summative online assessment where marks contributed towards the course grade, but significantly lower in formative online assessment where marks did not contribute towards the course grade. In a different study, Holmes (2018) found that the introduction of weekly summative e-assessments led to a significant increase in virtual learning environment activity compared to the virtual learning environment activity in that module the previous year.

However, research into the benefit of incentivisation on student learning shows a more mixed picture. Some recent studies have found no benefit to making assignments summative in order to increase engagement. In fact, Haugan et al. (2017) found that students spent more time studying and performed better when regular summative assessments were made purely formative, whilst Hellem (2019) found no difference in student performance when mandatory assignments were removed. Indeed, Gibbs (2010) expressed concerns over the use of this kind of incentivisation and its perceived benefit on learning:

> Students can tackle assignments that are intended as learning activities so as to maximise the marks they obtain rather than maximising the learning achieved from engaging with the assignment. (p. 11)

On the other hand, Brown et al. (2014) claim that in-class quizzes are more effective if you make them count towards the course grade. Freeman et al. (2007) showed that the impact of incentivised engagement varies according to the student. For high-risk students (those most likely to fail or perform poorly), they found that incentivisation led to improved results ($p = .08$, Hedges' g $= .31$). However, it also led to significantly worse results ($p = .034$, Hedges' g $= .33$) for low-risk students (those performing above average).

There is a wealth of research that has shown that extrinsic rewards can reduce intrinsic motivation (e.g. Deci, 1971; Deci et al., 2001). This is particularly the case where the behaviour is perceived to be controlled (Deci et al., 1999). In controlled behaviour, the cause of the behaviour is external to the person, e.g. when the reward for compliant behaviour is perceived by the individual as necessary, such as the award of grades. In contrast, autonomous behaviour is where the cause of the behaviour is internal to the person, where the behaviour is chosen and volitional.

Vansteenkiste et al. (2009) found that compared to someone who has low or no motivation, having high autonomous motivation (without controlled motivation) leads to significant learning, but having high controlled motivation (without autonomous motivation) does not. This suggests that if the goal of incentivisation is to improve the learning of students who are not otherwise motivated to study appropriately, then it may not be effective: poor-quality motivation may be no better than low-quantity motivation. On the other hand, as noted earlier, incentivisation has been shown to lead to improved results for students most likely to fail or perform poorly (Freeman et al., 2007).

2.2. Perusall in the literature

The literature proposes two complementary theories as to why a social annotation platform such as Perusall might improve student exam performance. The first is that the automated scoring system incentivises better study habits, deeper engagement with the pre-class reading and higher-level cognitive skills, which all contribute to higher learning gains (Cecchinato & Foschi, 2020; Miller et al., 2018). The second is that the social interaction through annotating and commenting in a shared environment motivates and improves student participation and learning (Itow, 2020; Murphy, 2021). This second theory aligns with constructivist theory that it is discussion with others and collaborative learning that leads to successful learning (Miller et al., 2016; Theodosiou & Corbin, 2020).

There is evidence that Perusall encourages more student reading and engagement at a higher cognitive level. Perusall's creators (Miller et al., 2018) show that 80% of students make it through 95% of the reading set for them. Adams and Wilson (2020) report that the number of annotations increased by 30% over the course of a semester and that the number, length or complexity of the reading had no impact on engagement. However, their study didn't investigate the quality of those annotations or the impact on learning. However, McFarlin (2020) report that a consequence of using Perusall is that students are better prepared for class discussion through more active reading and greater levels of critical thinking. Lee and Yeong (2018) noted that student annotations varied in complexity from simply defining terms through to explanations, links to other papers and concept maps, but that there was a low number of these higher cognitive level comments. Miller et al. (2016) suggests that the more desirable higher-level comments can be encouraged by seeding the discussion with some higher-level comments from a previous cohort. Overall, the consensus is that the use of Perusall enhances student engagement with the text but there is less consensus regarding the quality of the engagement.

In terms of collaborative learning, Adams and Wilson (2020) report that peer-to-peer interactions increased by 40% over the course of the semester using Perusall. They interpreted this as evidence of community growth but did not measure the quality of these interactions or the extent to which online interaction can complement or substitute for in-person interaction. In McFarlin (2020) the majority of students perceived Perusall to help improve connectedness with peers and instructors but in Theodosiou and Corbin (2020) the opposite was true. It seems more research is required on Perusall's impact in building community in a fully online setting.

Several studies have investigated the impact of social annotation on exam performance. Miller et al. (2016) reports students making higher quality explanations make more gains in conceptual understanding and consequently do better in exam performance. Miller et al. (2018) and Walker (2019) both note, when controlling for prior attainment, that students do better in a range of assignments in cohorts that use Perusall compared to cohorts that don't. In Greiger and Leontyev (2021) students were neutral in their self-perception of whether Perusall was beneficial in enhancing their understanding.

The validity of the automated scoring system has been investigated by Cecchinato and Foschi (2020). The correlation between Perusall marks and teacher marks was .58 ($p < .001$) showing moderate agreement. Walker (2019) also checked Perusall marks and found no need to adjust any marks. Student perception of the automated assessment is generally positive, but they believe it needs supervision to ensure validity (Cecchinato & Foschi, 2020).

Instructors report many positive features of Perusall. Bharath and Brownson (2021) highlight that it is user-friendly and the automated grading keeps students accountable. This view is also held by students (Greiger & Leontyev, 2021). Further, Clarke (2021) notes that it might draw out quieter students and allows instructors to identify struggling students. In addition, its creators (Miller et al., 2018) point out that it includes many features common on social media platforms (e.g. avatars, the ability to tag others, upvoting) and its timely feedback helps students and instructors to use time effectively. There are some caveats. For example, it does require internet access, the annotations can't be exported so some students keep additional notes offline and, in large classes, there are simply too many annotations for instructors to respond to (Clarke, 2021). The platform is free to use if, as in our study, the course materials are created and uploaded by the instructor but if the instructor wishes to use a textbook, the student must purchase the eBook through the Perusall platform which adds additional cost and restricts the instructor to the textbooks available in the Perusall catalogue.

Overall, there are relatively few papers which have studied Perusall in depth because it is a relatively new tool. There are, for example, no studies investigating its use in the Australian education system or in the teaching of statistics. However, the number of papers increased rapidly as the Covid-19 pandemic encouraged greater use of online tools. The literature supports the view that Perusall encourages engagement with pre-class reading and its use positively impacts performance in assignments. However, further work is required to understand why it has an effect and what factors, such as class size or the subject of the class, enhance or diminish that effect. In most cases (Lee & Yeong, 2018; Miller et al., 2016; Miller et al., 2018; Theodosiou & Corbin, 2020; Walker, 2019) student annotations are worth a small percentage of the final grade (typically 5-15%). No study has directly investigated whether this incentivisation has any impact on student engagement with Perusall or final grades.

3. Methods

3.1. Participants

The participants of this study were undergraduate students enrolled in the second-year course *Engineering Probability and Statistics* at an Australian university, in the second half (Semester 2) of 2020 over a 12-week semester. The majority of students (95%) were enrolled in Engineering degrees and taking *Engineering Probability and Statistics* as a core course for their degree. The remaining students were enrolled in *Engineering Probability and Statistics* as an elective or as a core course for their chosen major. In total, there were 74 students enrolled in the course, 54 of whom were enrolled at a Melbourne-based campus, and 20 of whom were enrolled at a regional campus. Of the 74 students enrolled in the course, 70 students completed the course (i.e. attempted the exam).

3.2. Course background

The course comprised a mix of weekly asynchronous and synchronous activities and was taught over a 12-week semester. During the first part of the semester (Weeks 1-6), topics covered include probability, probability distributions, location and spread, functions of random variables, and queues. During the second part of the semester (Weeks 7-11), topics covered include histograms and estimators followed by inferential statistics. The main asynchronous activity was a set of weekly readings covering main concepts, which were sometimes accompanied by short videos. There was also a weekly 2-hour computer lab where students would work through questions related to that week's topic(s). Due to the COVID-19 pandemic, all computer labs were held online via Zoom. Assessment was structured as follows: five fortnightly quizzes (10% total), four assignments with questions similar in nature to computer lab questions (40% total), and an end-of-semester exam (50%).

The weekly readings were provided via the Perusall platform so that students could interact with teaching staff and each other while engaging with the readings. Instructions about how to use Perusall were provided via a short video, which explained how to log in, how to make and respond to annotations, and how to navigate the platform. The video also explained that initially, students' Perusall activity would not contribute to their grades, but that it would later on in the semester. Student annotations were anonymous, but visible to the entire cohort. Each week, students would receive a Perusall 'score', which was calculated as a function of six components designed to measure engagement: annotations, opening the readings, reading completeness, active reading, getting responses on comments, and upvoting other students' comments. (More information about how Perusall scores are calculated can be found at https://www.perusall.com/hubfs/downloads/scoring-details.pdf.) As such, we use Perusall scores as a proxy for engagement with Perusall, and therefore engagement with the weekly readings.

For the second two assignments, 5 out of the 50 marks available on each assignment were based on students' Perusall scores related to the relevant weeks' readings. This meant that Perusall scores from Weeks 1–6 did not contribute to assignment marks, scores from Weeks 7–8 contributed 5 out of 50 marks for Assignment 3, scores from Weeks 9–10 contributed 5 out of 50 marksfor Assignment 4, and scores from Week 11 did not contribute to assignment marks. There were no readings in Week 12. During the mid-semester break, which was between Week 6 and Week 7, students were informed that their Perusall scores from Weeks 7 & 8, and 9 & 10, would comprise some of the marks for Assignments 3 and 4 respectively. We therefore consider the introduction of Perusall scores contributing to student marks during Weeks 7–10 as incentivisation, and the main intervention in this study.

3.3. Data collection

Student performance data from the assignments and final exam, and Perusall scores, are analysed. To allow for a meaningful comparison in assignment performance before and after the intervention, students who completed the course were included in the analysis, resulting in $n = 70$ students. Ethical approval to carry out the research study with a waiver of consent was obtained from the university human ethics research committee, RN HEC21427.

3.4. Data analysis

Visual inspection of a scatter plot followed by k-means clustering (MacQueen, 1967) was used to identify clusters of students with common characteristics during the pre-intervention phase of Weeks 1-6. The Chi-squared test of independence (see, e.g. McHugh, 2013) was used to test for an association between cluster and campus, and between cluster and the effect of the intervention on both Perusall engagement and student performance. Linear mixed effects models were used to compare both Perusall engagement and student performance over time. All Perusall scores and assignment and exam marks were converted to percentages for ease of comparison. Since marks as a percentage are bounded between 0 and 100, there is potential violation of the normality assumption for some models. Therefore, all the analysis was repeated using beta regression models to assess whether the results were sensitive to the choice of model. Results from both methods were similar throughout, suggesting no concerns regarding validity of the linear mixed effects model results, which have the advantage of ease of interpretability. Statistical analyses were carried out using R version 4.1.0 (R core team, 2021).

4. Results

We start by exploring characteristics of the student cohort (Section 4.1). Next, the impact of the intervention is evaluated by comparing Perusall scores with and without incentivisation (Section 4.2). In Section 4.3, student performance before and after the intervention is compared. The analysis in Sections 4.2 and 4.3 is presented for the cohort as a whole, as well as for each cluster (as identified in Section 4.1) separately. In Section 4.4, we explore whether there is an association between changes in Perusall engagement after the intervention, and changes in student performance, and whether this differs by cluster.

4.1. Characteristics of the student cohort

To inform the analysis carried out in the following sections, we explored whether students naturally fall into different groups with regard to their early Perusall engagement and assignment performance. This would then allow us to observe whether different groups of students tended to respond differently to incentivisation. Perusall scores from Weeks 1–6 were averaged to provide a summary 'before intervention' Perusall score. Similarly, marks from Assignments 1 and 2 were averaged to provide a summary 'before intervention' initial assignment performance measure. As a first step, we created a scatterplot of early assignment performance versus early Perusall scores (see Figure 1) and identified that there appeared to be three main groups, or clusters, of students as follows:

- Cluster 1: Students with relatively low assignment marks before the intervention. Most of these students also had relatively low Perusall engagement before the intervention
- Cluster 2: Students with relatively high assignment marks and Perusall engagement before the intervention
- Cluster 3: Students with relatively high assignment marks and relatively low Perusall engagement before the intervention

In order to objectively determine the boundaries of each group, a k-means clustering analysis, using initial assignment performance and initial Perusall engagement as input variables, was carried out. The k-means clustering algorithm works by deciding on the number of clusters, k, and then randomly assigning k central points (i.e. *centroids*). Observations are then assigned to clusters that have the closest centroid. Based on these allocations, the k centroids are updated, and observations re-allocated based on these updated centroids, with this process being repeated until either an optimal solution or a chosen maximum number of iterations is reached. The results of the k-means cluster procedure (with $k = 3$) are shown in Figure 1.

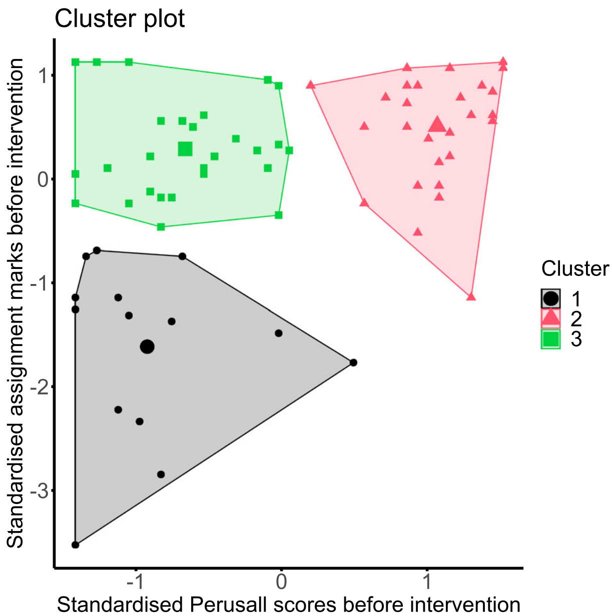

Figure 1. *k*-means cluster analysis results. The larger symbol in each cluster represents the mean (i.e. *centroid*) of that cluster.

Table 1. Two-way table showing the number of students from each campus separated by cluster.

Campus	Cluster 1	Cluster 2	Cluster 3	Total
Regional Campus	1	15	3	19
Melbourne Campus	13	14	24	51
Total	14	29	27	70

Of interest is that, as shown in Table 1, of the 14 students in Cluster 1, only one student was from the regional campus, with the other 13 students being from the Melbourne campus. Most of the 19 students from the regional campus were in Cluster 2 (15), while the highest number of Melbourne students were in Cluster 3 (24).

A Chi-squared test of independence was carried out to determine whether there is an association between campus and cluster. The relationship between these variables was significant, $\chi^2(2, N = 70) = 15.2$, $p < .001$. The general difference in characteristic of the student cohort between campus is useful to observe, because, for example, if it appears that some cluster(s) do or do not tend to benefit from incentivisation, campus-specific adjustments can be made.

4.2. Comparison of Perusall scores with and without incentivisation

In order to compare Perusall engagement with and without incentivisation, Perusall scores have been summarised into three time points:

- Time 1: Average Perusall score from Weeks 1–6 (no incentivisation)

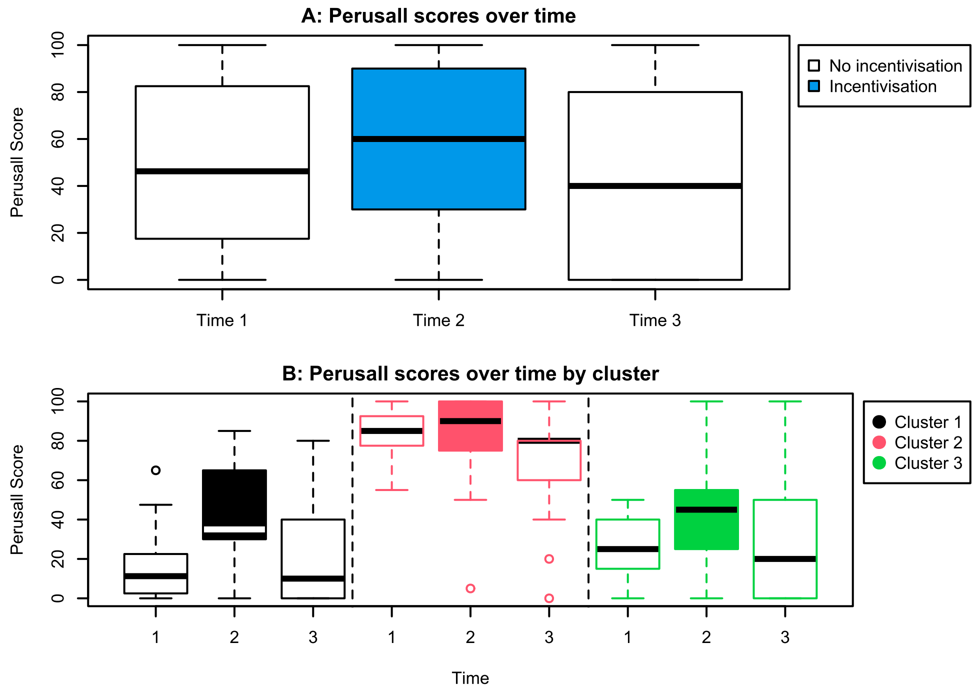

Figure 2. (a) Perusall scores over time for the entire cohort. (b) Perusall scores over time separated by cluster.

Table 2. Linear mixed-effects model to estimate average Perusall scores over time. The Intercept (Time 1) estimate refers to the average Perusall score at Time 1. The Time 2 and Time 3 estimates are the changes in average Perusall score from Time 1.

	Estimate (SE)	*t*-value (df)	*p*-value
Intercept (Time 1)	48.29(4.07)	11.85(138)	$<.001$
Time 2	11.5 (3.42)	3.37(138)	.001
Time 3	−4.29(3.42)	−1.25(138)	.212

- Time 2: Average Perusall score from Weeks 7–10 (incentivisation)
- Time 3: Perusall score from Week 11 (no incentivisation)

Students' Perusall scores at each time point are shown in Figure 2.

From Figure 2, we can observe that in terms of Perusall engagement, students in Cluster 1 greatly benefit from incentivising, students in Cluster 2 are already motivated and therefore gain little benefit from incentivisation, and students in Cluster 3 benefit marginally from incentivising.

To determine whether Perusall engagement was significantly different depending on whether incentivisation was present, we firstly consider overall Perusall scores as shown in Figure 2(a). Table 2 shows the results of a linear mixed-effects model, a repeated-measures analysis, with Perusall score as the response variable, and including Time as a fixed effect and student ID as the random effect.

Starting at Time 1 and with no incentivisation, the mean Perusall score was 48.29% (95% CI [40.23, 56.34]). When incentivisation was present during Time 2, Perusall scores were 11.5 (95% CI [4.75, 18.25]) percentage points higher, on average, and this difference was significant ($p = .001$). At Time 3, when incentivisation was no longer present, Perusall scores were 4.29 percentage points lower than at Time 1 and this difference was not significant. These results allow us to conclude that overall, student engagement with the weekly readings was significantly higher when incentivisation was present.

In order to compare Perusall scores over time for each cluster separately, as shown in Figure 2(b), a linear mixed-effects model was carried out as specified previously but for each cluster separately. The results for Clusters 1 and 3 were similar to the overall results in that, when comparing with Time 1, average Perusall scores were significantly higher at Time 2 ($p = .002$ and $p = .004$ respectively) but not at Time 3. By contrast, the results for Cluster 2 showed that, on average, there was no significant difference in Perusall engagement between Time 1 and Time 2. Average Perusall scores at Time 3 were 14.91 (95% CI [−23.92, – 5.90]) percentage points lower at Time 3 than at Time 1 and this difference was significant ($p = .002$). However, the beta model analysis found that this difference was not significant, this being the only difference found via the sensitivity analysis throughout the study. These results show that for the two clusters that had a low average Perusall score at Time 1 (16.79%, 95% CI [3.12, 30.45] and 25.65%, 95% CI [15.50, 35.79] for Clusters 1 and 3 respectively), incentivisation of Perusall engagement had a significant impact, with percentage point increases of 23.93 (95% CI [9.30, 38.56]) and 18.24 (95% CI [5.99, 30.49]) respectively. As Cluster 2 had a starting average Perusall score of 84.57% (95% CI [76.98, 92.16]), a ceiling effect is likely to have occurred such that incentivisation had little to no effect on the already high level of Perusall engagement.

4.3. Comparison of student performance before and after incentivisation

In order to compare student performance before and after incentivisation, we summarise student assessment into three time points:

- Time 1: Average mark on Assignments 1 and 2 (pre incentivisation)
- Time 2: Average mark on Assignments 3 and 4 (post incentivisation)
- Time 3: Exam mark (post incentivisation)

Of note here is that we consider Time 1 to be 'before' incentivisation, and Times 2 and 3 to be 'after' incentivisation. This is because introducing incentivisation half-way through the semester may not only have had an effect on student performance to the associated assignments (Assignments 3 and 4), but also on overall attainment as measured by the final exam. Student performance at each time point is shown in Figure 3.

Figure 3 indicates that in terms of performance, students in Cluster 1 greatly benefit from incentivising, students in Cluster 2 are already performing well and therefore gain little benefit from incentivisation, and students in Cluster 3 do not benefit from incentivising.

We again consider the class as a whole, when comparing student performance over time, by carrying out a linear mixed-effects model. Table 3 shows the results of this analysis, with

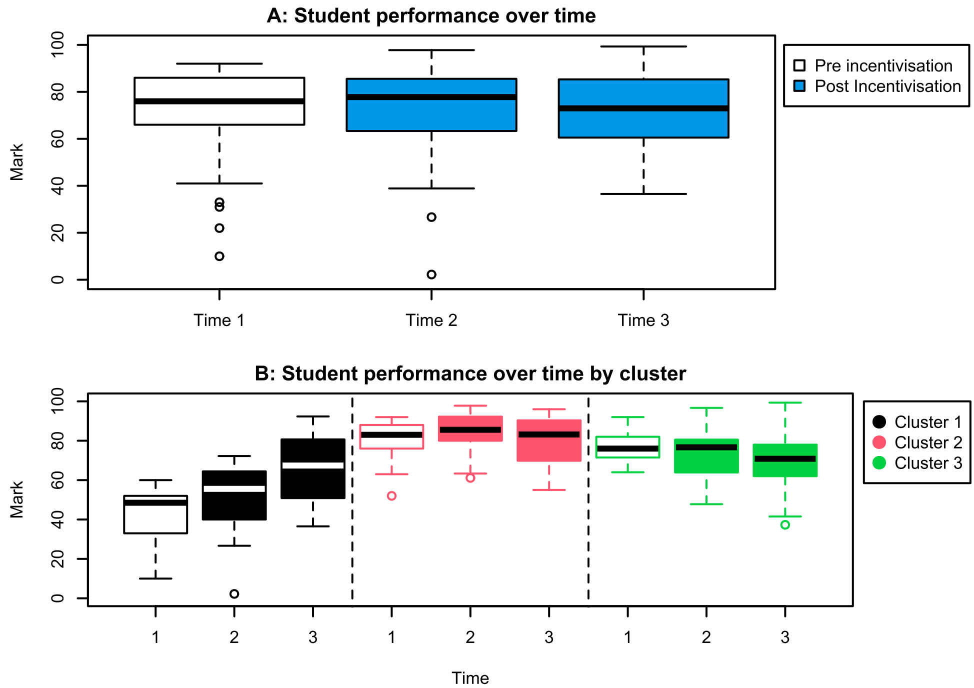

Figure 3. (a) Student performance over time for the entire cohort. (b) Student performance over time separated by cluster.

Table 3. Linear mixed-effects model to estimate average student performance over time. The Intercept (Time 1) estimate refers to the average mark at Time 1. The Time 2 and Time 3 estimates are the changes in average mark from Time 1.

	Estimate (SE)	*t*-value (df)	*p*-value
Intercept (Time 1)	72.14 (2.06)	34.98 (138)	$< .001$
Time 2	0.70 (1.91)	0.37 (138)	.715
Time 3	0.74 (1.91)	0.39 (138)	.698

student mark as the response variable, and including Time as a fixed effect and student ID as the random effect.

As expected, based on the boxplots shown in Figure 3(a), there was no significant difference on student performance before and after incentivisation when considering the group as a whole.

We again carry out a linear mixed-effects analysis for each cluster separately, this time looking at student performance over time. On average, students in Cluster 1 performed significantly better at Time 3 (mean increase = 21.90 (95% CI [11.95, 31.84]), $p < .001$) as compared with Time 1 (mean = 43.71%, 95% CI [34.30, 53.13]). While students in Cluster 1 also performed better at Time 2 (mean increase = 6.44), this difference was not significant. For Cluster 2, there were no significant differences in performance across time. Comparing from Time 1 (mean = 77.26%, 95% CI [72.49, 82.02]), average performance

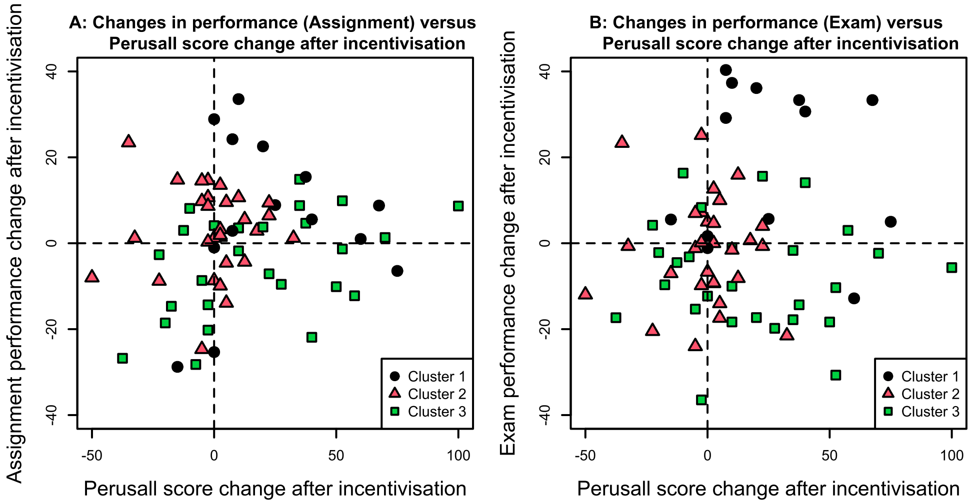

Figure 4. (a) Change in performance (Assignment) versus change in Perusall score after incentivisation. (b) Change in performance (Exam) versus change in Perusall score after incentivisation. Perusall scores after incentivisation do not include Week 11.

of Cluster 3 was lower at both Time 2 (mean decrease = 4.58) and Time 3, and the decrease at Time 3 was significant (mean decrease = 7.67, 95% CI [−13.25, – 2.08], $p = .008$).

4.4. Association between change in Perusall score and student performance after incentivisation

Figure 4 shows that after incentivisation, most students in Cluster 1 saw an increase in their Perusall scores. This coincided with an increase in their performance on both assignments and the exam, with most observations from Cluster 1 appearing in the top-right quadrant of both Figure 4(a,b). As discussed in the previous section, this increase in performance by Cluster 1 was significant. By contrast, observations from Clusters 2 and 3 are spread over all four quadrants of Figure 4(a,b) and there was not a significant increase in performance.

Table 4 shows the number of students from each cluster appearing and not appearing in the top-right quadrant of Figure 4(a,b) respectively. Although the majority (64.3%) of students in Cluster 1 appear in the top-right quadrant for both the assignment and the exam, this is not the case for Clusters 2 and 3. For the assignment, 37.9% and 33.3% of students from Clusters 2 and 3 respectively appear in the top-right quadrant. This is even less for the exam, with 20.7% and 11.1% respectively.

Chi-squared tests of independence were carried out to determine whether there is an association between appearance in top-right quadrant, and cluster. With regard to the exam, the relationship between these variables was significant, $\chi^2(2, N = 70) = 14.304$, $p < .001$, but not significant with regard to assignments, $\chi^2(2, N = 70) = 3.8897$, $p = .143$. These results support that incentivisation may have been of some benefit to students in Cluster 1 with regard to exam performance, while the same effect is not seen for students in Clusters 2 and 3.

Table 4. Two-way table showing the number of students with positive changes in both Perusall scores and performance (top-right quadrant) or not (other) by cluster, for both the assignment and the exam.

	Cluster 1	Cluster 2	Cluster 3	Total
	Assignment			
Top-right quadrant	9	11	9	29
Other	5	18	18	41
	Exam			
Top-right quadrant	9	6	3	18
Other	5	23	24	52

5. Discussion

Considering the cohort as a whole, the results indicate that incentivisation had a positive and significant effect on engagement with the weekly readings, but no significant effect on student performance. Considering the clusters separately, however, provides further insight. After the intervention, most students in Cluster 1 had a higher Perusall score and attained a higher mark on both assignments and the final exam. These increases were significant in most cases. Recalling that before the intervention, Cluster 1 had a relatively low level of engagement with Perusall coupled with relatively low assignment marks (i.e. considered a 'high-risk' group), the intervention seems to have been of benefit for Cluster 1. However, for Cluster 2, which already had relatively high levels of both engagement and attainment prior to the incentivisation, the intervention seems to have had no effect. In fact, engagement with Perusall decreased significantly for this cluster at Time 3 compared to Time 1, following the incentivisation. This is consistent with previous findings in the literature that extrinsic rewards can reduce intrinsic motivation (Deci et al., 2001), although other factors may also have had an impact here. For example, Time 3 occurs at the end of the semester, where enthusiasm and motivation is likely to be lower, and students were aware that the reading for this week would not be assessed directly in the final assignment. Clusters 1 and 3 did not show the same decrease in engagement with Perusall at Time 3 as compared with Time 1.

For Cluster 3, which had low engagement but high performance prior to the intervention, we observe a subsequent decrease in performance, particularly in the exam (see Figure 3(b)). The significant decrease in performance for Cluster 3 suggests that incentivisation may have had a negative impact on these students, although the causes of this decrease in performance are unclear. Freeman et al. (2007) found that incentivisation had a negative impact on high performing students. Considering that the students in Cluster 3 had low Perusall scores and high assignment marks before the incentivisation, it may be that these students' engagement was merely superficial, as a 'tick-box exercise' to gain marks, as noted by Gibbs (2010). Although Cluster 3 saw a significant increase in their engagement during the incentivisation period, it is possible that this engagement was superficial. The results suggest that while incentivisation may lead to a higher level of engagement, this higher engagement may (e.g. Cluster 1) or may not (e.g. Cluster 3) be sufficiently strong to lead to high end-of-semester performance. In addition, this may be dependent upon the student's approach to learning (deep, surface or strategic, see e.g. Entwistle (1988)) and motivational style.

It is widely acknowledged that reading mathematics, as opposed to general prose, is a specific skill that needs nurturing (Hodds et al., 2014) and not all tools designed to increase comprehension elicit the desired higher-level understanding (Alcock et al., 2015). Our results show that while students respond positively to Perusall in terms of their engagement, further work is needed to understand the cognitive processes it encourages, and whether these are helpful for mathematical thinking.

5.1. Limitations

This study was limited by its design, whereby incentivisation was applied to the whole cohort, halfway through the semester. Although this meant that students' initial assignment marks could be used as their own control for comparison, it also meant that assessing the effect of Perusall incentivisation on student performance was challenging. A randomised controlled trial would help to more clearly establish a cause-and-effect relationship, or lack thereof, between incentivisation and student performance, although this may prove difficult due to ethical considerations. Another consequence of introducing the intervention halfway through the semester is that topics covered before or after the intervention may have had different difficulty levels, which could potentially lead to confounding. However, two mitigating factors with regard to this potential limitation are that in this particular course, there was a mix of easier and harder topics covered in both parts of the semester, and students' initial assignment marks could be used as their own control for comparison.

6. Conclusions

The literature suggests that the use of Perusall combined with a mark incentive encourages deeper engagement with pre-class reading, which contributes to higher learning gains (Miller et al., 2018). The baseline period in our study shows that, without the additional incentive, there is high variability between students in their level of engagement with readings on the Perusall platform. However, our study demonstrates that incentivisation increases the engagement for the majority of students and, therefore, the incentivisation is an important component in Perusall's apparent success in the literature at improving engagement. In our study, the impact of Perusall, and the incentivisation of using Perusall, on learning gains is less clear, with it appearing to be beneficial for some students but not others. In particular, our findings show that in terms of student performance, incentivisation appeared to be of some benefit to high-risk students (those most likely to fail or perform poorly). For high-performing students, the effect appeared to be either neutral or negative, and this potentially depends on the motivational style of the student, i.e. whether or not incentivisation exacerbated a superficial level of engagement. Further work is required to better understand the link between a given student's approach to learning and motivational style, and how they are likely to respond to incentivised engagement, leading to recommendations about how it may be employed to support the learning of all students.

Acknowledgements

This research is derived from work discussed in books of abstract Shaker et al. (2022a) and Shaker et al. (2022b).

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Amanda J. Shaker http://orcid.org/0000-0002-4672-9142

References

Adams, B., & Wilson, N. S. (2020). Building community in asynchronous online higher education courses through collaborative annotation. *Journal of Educational Technology Systems*, *49*(2), 250–261. https://doi.org/10.1177/0047239520946422

Alcock, L., Hodds, M., Roy, S., & Inglis, M. (2015). Investigating and improving undergraduate proof comprehension. *Notices of the AMS*, *62*(7), 742–752.

Anderson, J. (2016). Student engagement and the learning incentive program: Evidence and applications. *Sensoria: A Journal of Mind, Brain & Culture*, *12*(1), 28. https://doi.org/10.7790/sa.v12i1.429

Beard, L. H. (2017). 'Incentivised reading'- using an online VLE to measure engagement and attainment in student learning. *International Journal for Innovation Education and Research*, *5*(11), 74. https://doi.org/10.31686/ijier.vol5.iss11.854

Bharath, D. M. N., & Brownson, S. (2021). Perusall: Read, connect, discuss!. *Journal of Public Affairs Education*, *27*(3), 372–375. https://doi.org/10.1080/15236803.2021.1929021

Brown, P. C., Roediger IIIH. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. Belknap Press of Harvard University Press.

Burke, G., Macan Bhaird, C., & O'Shea, A. (2013). The effect of a monitoring scheme on tutorial attendance and assignment submission. *International Journal of Mathematical Education in Science and Technology*, *44*(4), 545–553. https://doi.org/10.1080/0020739X.2012.756553

Cecchinato, G., & Foschi, L. C. (2020). Perusall: University learning-teaching innovation employing social annotation and machine learning. *Open and Interdisciplinary Journal of Technology, Culture and Education*, *15*(2), 45–67.

Clarke, A. J. (2021). Perusall: Social learning platform for Reading and annotating (perusall LLC, perusall.com). *Journal of Political Science Education*, *17*(1), 149–154. https://doi.org/10.1080/15512169.2019.1649151

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*(1), 105–115. https://doi.org/10.1037/h0030644

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627–668. https://doi.org/10.1037/0033-2909.125.6.627

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research*, *71*(1), 1–27. https://doi.org/10.3102/00346543071001001

Entwistle, N. (1988). *Styles of learning and teaching*. David Fulton Publishers.

Freeman, S., O'Connor, E., Parks, J. W., Cunningham, M., Hurley, D., Haak, D., Dirks, C., & Wenderoth, M. P. (2007). Prescribed active learning increases performance in introductory biology. *CBE—Life Sciences Education*, *6*(2), 132–139. https://doi.org/10.1187/cbe.06-09-0194

Gibbs, G. (2010). *Using assessment to support student learning*. Leeds Met Press.

Greiger, K., & Leontyev, A. (2021). Student-generated infographics for learning green chemistry and developing professional skills. *Journal of Chemical Education*, *98*(9), 2881–2891. https://doi.org/10.1021/acs.jchemed.1c00446

Haugan, J., Lysebo, M., & Lauvas, P. (2017). Mandatory coursework assignments can be, and should be, eliminated!. *European Journal of Engineering Education*, *42*, 1408–1421.

Hellem, V. (2019). *The Effect of Mandatory Assignments on Students' Learning Outcome and Motivation in Introductory Programming Courses* [Master's thesis]. Norwegian University of Science and Technology.

Hodds, M., Alcock, L., & Inglis, M. (2014). Self-explanation training improves proof comprehension. *Journal for Research in Mathematics Education*, *45*(1), 62–101. https://doi.org/10.5951/jresemath educ.45.1.0062

Holmes, N. (2018). Engaging with assessment: Increasing student engagement through continuous assessment. *Active Learning in Higher Education*, *19*(1), 23–34. https://doi.org/10.1177/146978741 7723230

Indiana University Center for Postsecondary Research. (2002). *From Promise to Progress: How Colleges and Universities are Using Student Engagement Results to Improve Collegiate Quality*.

Itow, R. C. (2020). Fostering valuable learning experiences by transforming current teaching practices: Practical pedagogical approaches from online practitioners. *Information and Learning Sciences*, *121*(5), 443–452. https://doi.org/10.1108/ILS-04-2020-0106

Lee, S. C., & Yeong, F. M. (2018). Fostering student engagement using online, collaborative reading assignments mediated by Perusall. *The Asia Pacific Scholar*, *3*(3), 46–48. https://doi.org/10.29060/TAPS.2018-3-3/PV2000

Macfarlane, B., & Tomlinson, M. (2017). Critical and alternative perspectives on student engagement. *Higher Education Policy*, *30*(1), 1–4. https://doi.org/10.1057/s41307-016-0026-4

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*(14), 281–297.

McFarlin, T. J. (2020). *Teaching property law: Using open-source, collaborative online reading to teach property*. Saint Louis University Law Journal.

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia Medica.* 23(2):143–149. https://doi.org/10.11613/BM.2013.018

Miller, K., Lukoff, B., King, G., & Mazur, E. (2018). Use of a social annotation platform for pre-class reading assignments in a flipped introductory physics class. *Frontiers in Education*, *3*, 8. https://doi.org/10.3389/feduc.2018.00008

Miller, K., Zyto, S., Karger, D., Yoo, J., & Mazur, E. (2016). Analysis of student engagement in an online annotation system in the context of a flipped introductory physics class. *Physical Review Physics Education Research*, *12*(2), 020143. https://doi.org/10.1103/PhysRevPhysEducRes.12.020143

Murphy, J. A. (2021). Collaborative annotation: Tools for enhancing learning and scholarly communication. *Serials Review*, *47*(3-4), 157–162. doi:10.1080/00987913.2021.1986917

R Core Team. (2021). *R: A language and environment for statistical computing [Computer program]*. R Foundation for Statistical Computing. https://www.R-project.org/.

Shaker, A., Brignell, C., & Pugh, M. (2022a). Incentivisation and its effect on engagement and student performance, In Book of Abstracts, CETL-MSOR 2022, Abertay University, Dundee, UK, pp. 29–30. Link: https://www.conventiondundeeandangus.co.uk/uploads/tinymce/DR/2022/CETL-MSOR2022/Book%20of%20Abstracts%20FINAL.pdf.

Shaker, A., Brignell, C., & Pugh, M. (2022b). Incentivisation and its effect on engagement and student performance, In conference booklet, AustMS 2022, University of New South Wales, Sydney, Australia, p. 137. Link: https://web.maths.unsw.edu.au/ pinhas-grossman/booklet_nov29.pdf.

Theodosiou, N. A., & Corbin, J. D. (2020). Redesign your in-person course for online: Creating connections and promoting engagement for better learning. *Ecology and Evolution*, *10*(22), 12561–12572. https://doi.org/10.1002/ece3.6844

Trowler, V. (2010). *Student engagement literature review*. Higher Education Academy.

Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K., & Lens, W. (2009). Motivational profiles from a self-determination perspective: The quality of motivation matters. *Journal of Educational Psychology*, *101*(3), 671–688. https://doi.org/10.1037/a0015083

Walker, A. S. (2019). Perusall: Harnessing AI robo-tools and writing analytics to improve student learning and increase instructor efficiency. *The Journal of Writing Analytics*, *3*(1), 227–263. https://doi.org/10.37514/JWA-J.2019.3.1.11

Weems, G. (1998). The impact of homework collection on performance in intermediate algebra. *Research and Teaching in Developmental Education*, *15*, 21–25.

Wimpenny, K., & Savin-Baden, M. (2013). Alienation. Agency and Authenticity: A Synthesis of the Literature on Student Engagement. *Teaching in Higher Education*, *18*(3), 311–326.