# Using Gamification to Normalise Fine-grained Assessment of Classroom Engagement.

David N. Milne
*Faculty of Engineering and IT, University of Technology, Sydney, Australia*
*Corresponding Author Email: David.Milne@uts.edu.au*

## ABSTRACT

### CONTEXT

Engagement is an integral component of successful learning. Fine-grained measurement and formal assessment of engagement may help promote it by highlighting it as something that students are required to demonstrate. However there is a strong risk that assessment of engagement could undermine itself if it lacks robustness (i.e. if it measures and promotes a poor proxy of engagement) or creates undue stress (which inhibits engagement).

### PURPOSE OR GOAL

This paper introduces and evaluates Gameface, an application that we have designed to be deployed in tutorials and laboratory classes to measure and promote engagement. Gameface aims to assess behavioural and cognitive engagement through fine-grained, comprehensive measurement of classroom activity. It aims to do so transparently without creating undue stress by promoting collaboration and employing gamification to maintain a playful tone.

### APPROACH OR METHODOLOGY/METHODS

Gameface was deployed (and used for as a component of formal assessment) in a large introductory programming subject. Students were surveyed in the middle and end of the teaching session about whether they endorsed Gameface being adopted more broadly, and why. The survey responses were analysed both quantitively (to measure the extent of endorsement) and qualitatively (to uncover the reasons why).

### ACTUAL OR ANTICIPATED OUTCOMES

In the middle of the teaching session, 63% of respondents felt that Gameface would make other subjects better, while 17% felt it would make them worse. By the end of the teaching session, 74% felt it would make other subjects better, while 9% felt it would make them worse. The difference between the two surveys was statistically significant, indicating that changes made to Gameface alleviated many of the concerns that students initially had. The thematic analysis showed that Gameface was perceived by students as improving behavioural engagement (e.g. encouraging students to prepare well for class and be active participants during class), cognitive engagement (e.g. helping students stay switched on and focused), emotional engagement (e.g. maintaining a playfully competitive tone) and social engagement (helping students work with and learn from their peers). However it also revealed that the use of Gameface for assessment may have undermined engagement for some students, particularly prior to some key design changes.

### CONCLUSIONS/RECOMMENDATIONS/SUMMARY

This study provides an example of how to apply gamification to achieve fine-grained measurement and assessment of classroom engagement without losing student endorsement.

### KEYWORDS

engagement, gamification, assessment

# Introduction

Students with high engagement are more likely to have good learning outcomes, which in-turn leads to greater engagement (Marcum, 2000). By measuring engagement, teachers and institutions can make data driven decisions to support this virtuous cycle (Bond et al., 2020). Many definitions of engagement have been proposed, but they generally converge on three key dimensions: cognitive, behavioural and emotional/affective (Mandernach, 2015). Recent studies have added social engagement as a fourth dimension (Wang et al., 2016).

Students align their efforts to assessment (Gibbs, 2006), so one way to promote engagement may be to formally assess it. In an era where generative AI is undermining our ability to verify learning outcomes (Cotton et al., 2023), assessing engagement may help us verify that learning has been undertaken. However there is a high risk that assessment of engagement will undermine itself due to washback (Leber et al., 2017). It must be robust and comprehensive to avoid measuring (and promoting) a poor proxy rather than the real thing. Additionally, assessing it may promote stress, which directly undermines engagement (Alonso-Tapia et al., 2023).

Enter Gameface, a system that we have designed and developed to promote engagement within the classroom. Gameface enables assessment of engagement through fine-grained, comprehensive measurement of classroom activity. It aims to do so without creating undue stress by promoting collaboration and employing gamification to maintain a playful tone.

This paper describes an evaluation in which Gameface was deployed (and used for formal assessment) in a large introductory programming subject. Gameface received strong endorsement from students and was perceived as improving markers of engagement such as enjoyment, focus, collaboration, and willingness to expend effort.

# Gameface

Gameface is a web application that can be accessed via any device with a web browser. At the beginning of class, the tutor displays Gameface on the main projector screen (Figure 1b). Students can scan the QR code shown there to access the player screen (Figure 1a) on their own devices. This screen allows students to view their scores, find groupmates, and participate in quizzes and tasks. The tutor can scan the same QR code to access the remote control (Figure 1c) for controlling the room and accessing sensitive information such as scores, choices, and private notes about the attendees.

Please note that all names and faces in Figure 1 (and any that follow) are fictional, and the avatars have been generated by AI.
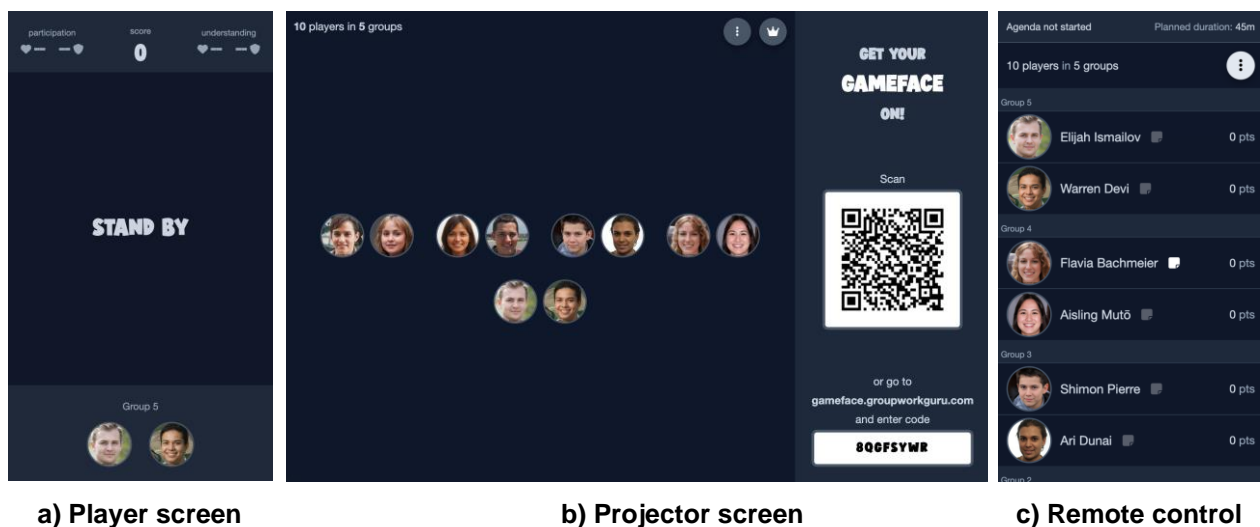


a) Player screen        b) Projector screen        c) Remote control

**Figure 1: The Gameface web application**

## Eliciting engagement

Gameface provides three key activities for engaging students: *quizzes*, *volunteer tasks*, and *class-wide tasks*.



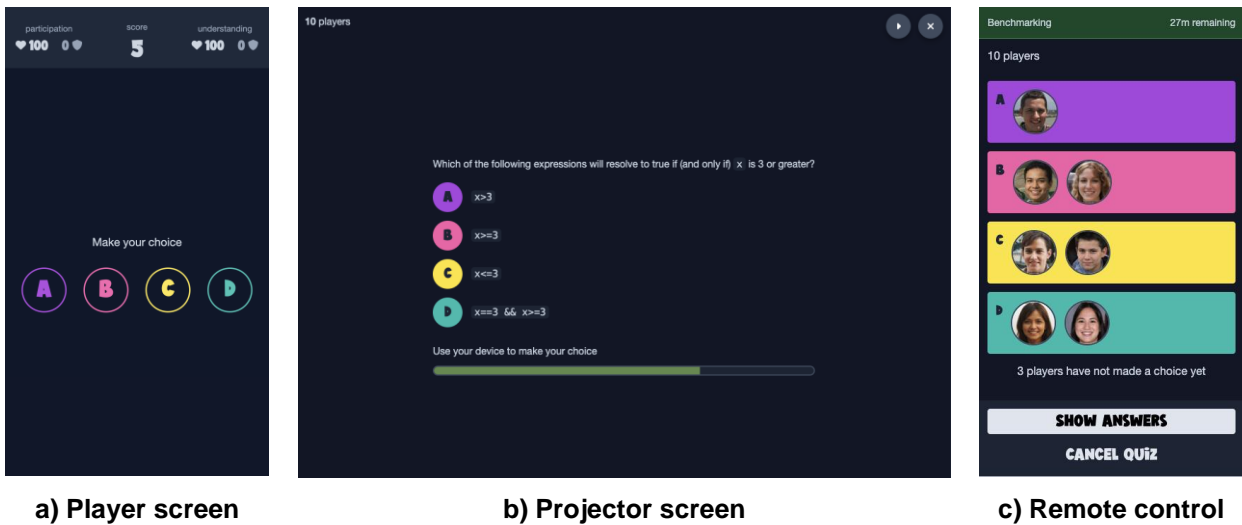a) Player screen          b) Projector screen          c) Remote control

**Figure 2: Running quizzes in Gameface**

Figure 2 shows how quizzes function in Gameface. The projector screen displays the question, possible answers, and a real-time progress bar indicating the proportion of students who have answered. The tutor can use this data to choose when to close the question and reveal the correct answer. At this point students' screens will reveal whether they answered correctly, and any points earned. The projector screen shows the proportion of students who chose each option, so that the tutor can address any misconceptions (i.e. if a significant proportion of the class has made an incorrect choice) before proceeding. The remote control additionally shows the choice made by each individual student.
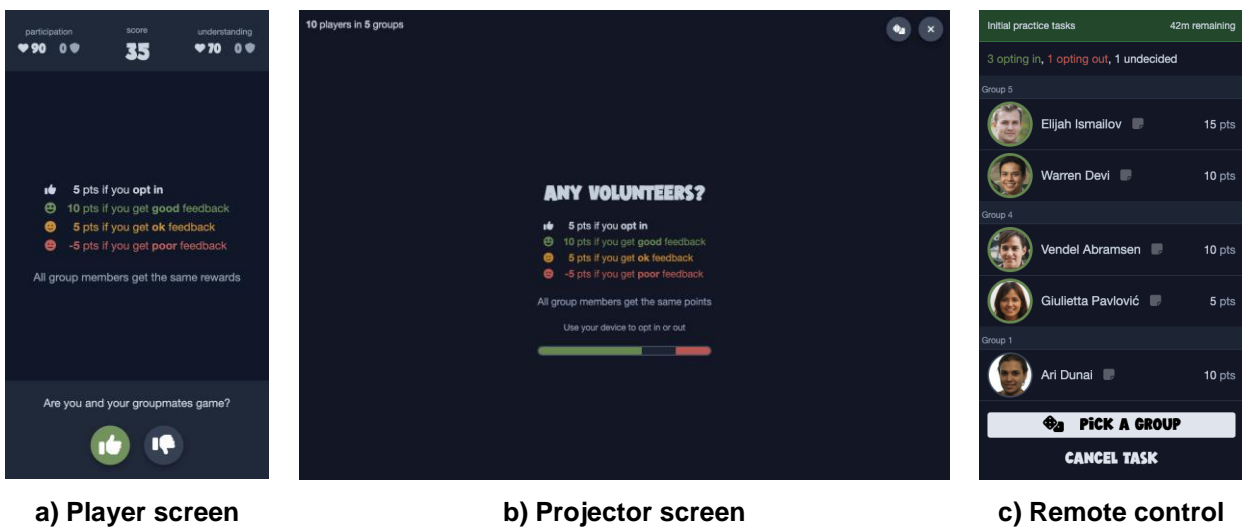


a) Player screen          b) Projector screen          c) Remote control

**Figure 3: Recruiting volunteers in Gameface**

A *volunteer task* is a lightweight mechanism that tutors can engage whenever they need a volunteer to answer a question or make some other contribution to the class (Figure 3). When a tutor launches a volunteer task (which they can do at any time) they specify whether they require an individual or a group to volunteer, and the points offered as a reward. When a task is launched, the projector screen updates to display the available rewards, and a progress bar of the proportion of the class who has opted in or out of the task. Each student uses their own device to decide whether to opt in or out. The remote-control updates in real time to show the

decisions made by each student. Once the tutor has granted everyone enough time to respond they can move on, at which point an individual or group will be selected by the system and highlighted in the projector screen and the remote control. The tutor can then hear the selected individual or group out and confidentially score the contribution as *poor*, *ok* or *good* (in addition to any verbal feedback). The task closes once feedback is provided.
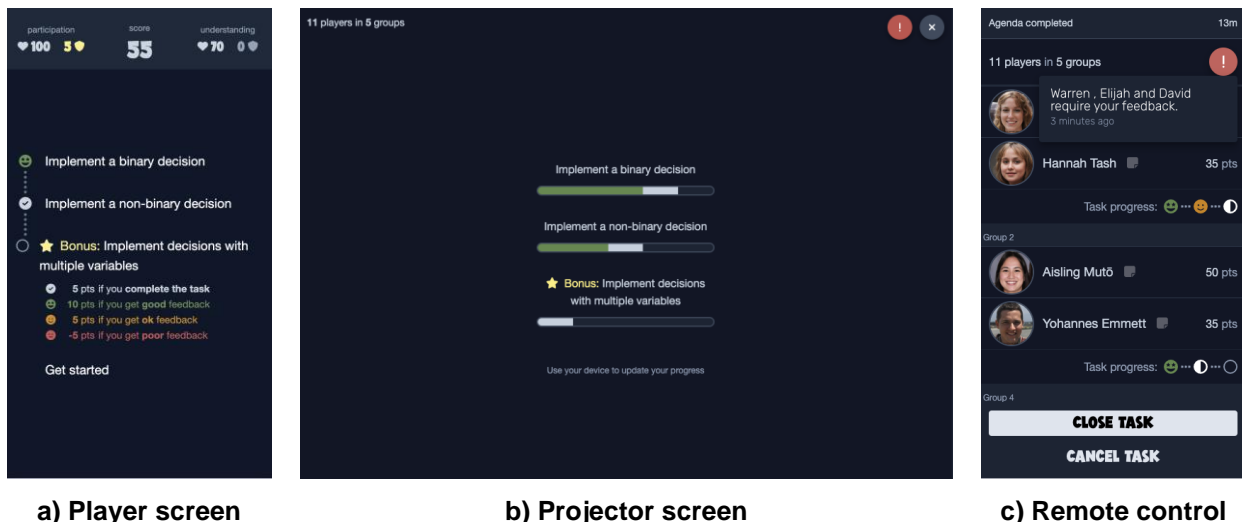


**a) Player screen**          **b) Projector screen**          **c) Remote control**

**Figure 4: Conducting class-wide tasks in Gameface**

*Class-wide tasks* allow the tutor to specify a sequence of activities for all students in the class to complete, either individually or in groups. As shown in Figure 4, once a class-wide task is launched each student can use their device to keep the tutor updated on their progress. The projector screen shows the proportion of the class that is attempting (in white) or has completed (in green) each activity. The remote control allows the tutor to view the progress and assess the contributions of each individual or group. Whenever students complete an activity, the remote control provides a notification that allows the tutor to locate the student(s), verify that the activity has been completed, and provide a rating of either *poor*, *ok* or *good* (in addition to any verbal feedback they might provide). The idea here is not for Gameface to dictate what these tasks are or how they are completed. Instead it is task-agnostic and leaves the tutor to scaffold tasks and provide any necessary resources as they normally would. Gameface merely provides lightweight means of tracking progress and providing/capturing quantitative feedback.

Volunteer tasks can be launched on demand, but quizzes and class-wide tasks require planning. Tutors and coordinators can create lesson plans in which each activity is given a time estimate and explanatory notes, and can refer to them on their remote control at any time. These lesson plans are particularly useful for maintaining consistency across teaching teams in larger subjects.

## Promoting collaboration

All the activities described above can be completed individually or in groups. Gameface allows tutors to allocate students to groups manually, or automatically using one of the following three strategies: a) the system randomly allocates students to groups, to help form new connections; b) the system ensures that members in each group have similar scores, to allow high performing students to excel together and for more focus to be given to those who are struggling; or c) the system ensures that each group has at least one member with a high score, to maximise peer learning. Students can view their allocated groupmates on their devices at any time.

## Assessing behavioural and cognitive engagement

Gameface provides measures of behavioural and cognitive engagement that are intended to be used for formal assessment.

Behavioural engagement (or *participation* in the figures above) is measured as the extent that a student capitalises on the opportunities they are given to engage. It is based on points earned from answering questions, opting in to volunteer tasks, and completing class-wide tasks, as a percentage of points offered. If a student answers all quiz questions, opts in to all volunteer tasks, and completes all class-wide tasks, then their behavioural engagement would be 100.

Cognitive engagement (or *understanding* in the figures above) is measured as the extent that a student successfully demonstrates their knowledge and ability. It is based on the points earned from answering questions correctly and getting positive feedback from volunteer and class-wide tasks, as a percentage of points offered. If a student answers all quiz questions correctly and gets *good* feedback for all volunteer and class-wide tasks, then their cognitive engagement would be 100. It will drop whenever they answer a question incorrectly or receive *poor* or *ok* feedback.

These scores can be used to assess whether students have engaged in the learning process actively and thoughtfully but can generate a great deal of stress and anxiety if misused. Without adjustment, every mistake or missed opportunity translates directly to students' final grades, which sets an unfair expectation that they must know everything before coming to class.

To make assessment more forgiving, each score is split into health (which should be used for assessment) and armour (which should not). Tutors can offer bonus questions and tasks which do not count towards the total points being offered (i.e. the total that health is a percentage of) and consequently can be safely skipped without losing participation health, or done incorrectly without losing understanding health. When a student's participation or understanding health is full, any additional points that they earn from bonus items becomes armour. Lost points are removed from armour before they have any impact on health. Thus bonus items provide a risk-free chance to recover lost health, or to gain a buffer to prevent future losses. They also allow students to choose how they demonstrate engagement. For example, a student who is uncomfortable speaking in front of the class can avoid volunteer tasks if they do sufficiently well with quiz questions and class-wide tasks.

# Evaluation method

This research was approved by the UTS Human Research Ethics Committee (ETH24-9530). Participants were recruited from an introductory programming subject held in Autumn 2024. All 348 students enrolled in the subject were required to use Gameface for 10 weeks (i.e. from weeks 2-11) during their laboratory classes. The participation and understanding scores described previously were averaged over the teaching session as an assessment item that accounted for 30% of student's final grade.

Participants were invited to complete two anonymous feedback surveys; one at the end of week 4 and the other at the end of week 11. Both surveys asked *"Would other subjects be better or worse if they used Gameface?"* and *"Why do you think Gameface would have this effect?"*. The first question was answered on a five-point scale ranging from *significantly worse* to *significantly better*, and the second question was open-ended.

Updates were made to Gameface throughout the teaching session. By far the most significant change was the introduction of bonus activities and armour in week 5. As described previously, this made assessment more forgiving and allowed students to choose how they demonstrate engagement. More minor changes were to provide detailed breakdowns for how points were earned or lost (released in Week 5), and to bias the random selection of volunteers to prioritise students with low scores and avoid repeatedly selecting the same student (Week 8).

Statistical analysis was conducted using the SciPy Scientific Library for Python (version 1.4.1). Thematic analysis was conducted manually following an inductive thematic approach (Braun & Clarke, 2006). All responses were read in their entirety to develop an initial set of themes. Responses were then annotated with the relevant themes and the sentiment expressed towards each theme. Each response could match to multiple themes, and could reflect a positive sentiment towards one theme and a negative sentiment towards another.

# Results

Figure 5 illustrates whether participants endorse Gameface being used more broadly in other subjects. It also compares responses captured during the middle (i.e. at week 4) and the end (week 11) of the teaching session. The week 4 survey received 155 responses, or 45% of the cohort. The week 11 survey received 116 responses, or 33% of the cohort.
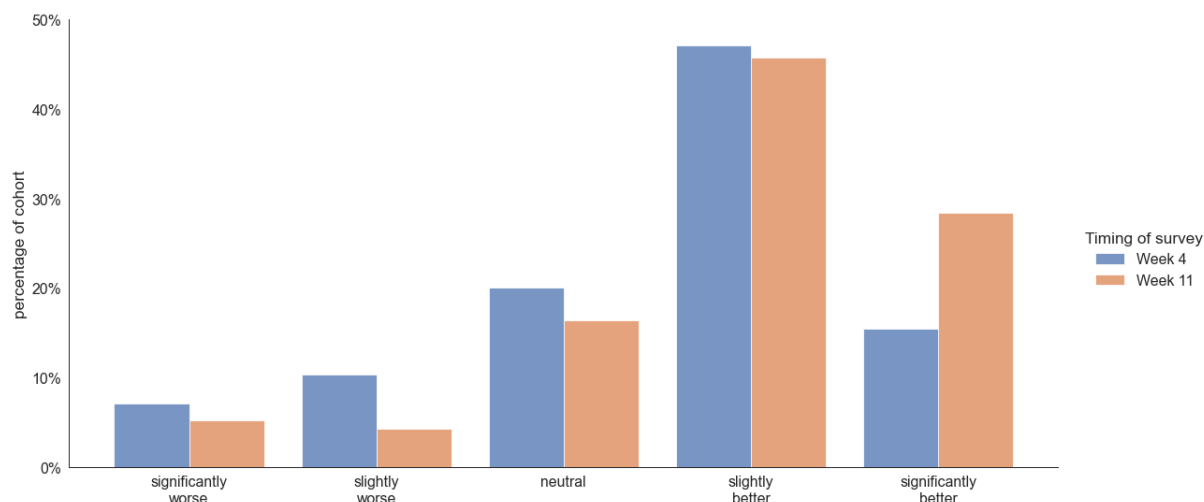


**Figure 5: Would other subjects be better or worse if they used Gameface?**

In week 4, 63% of respondents felt that Gameface would make other subjects better, while 17% felt it would make them worse. In week 11, 74% felt that Gameface would make other subjects better, while 9% felt it would make them worse. Endorsement of Gameface was significantly higher at the end of the teaching session (Mann–Whitney U = 7317.0, n1 = 155, n2 = 116, P = 0.003 one-tailed).

Table 1 summarises the themes that emerged when participants were asked to explain why they felt that Gameface would or would not improve other subjects.

**Table 1: Prevalence and sentiment of themes**

| theme | timing | mentions | | positive | | negative | |
|---|---|---|---|---|---|---|---|
| participation | week 4 | 60 | 39% | 57 | 95% | 3 | 5% |
| | week 11 | 47 | 41% | 46 | 98% | 1 | 2% |
| learning | week 4 | 38 | 25% | 32 | 84% | 6 | 16% |
| | week 11 | 38 | 33% | 35 | 92% | 3 | 8% |
| enjoyment | week 4 | 20 | 13% | 16 | 80% | 4 | 20% |
| | week 11 | 15 | 13% | 11 | 73% | 4 | 27% |
| collaboration | week 4 | 21 | 14% | 19 | 91% | 2 | 10% |
| | week 11 | 10 | 9% | 9 | 90% | 1 | 10% |
| fairness | week 4 | 23 | 15% | 6 | 26% | 17 | 74% |
| | week 11 | 7 | 6% | 6 | 86% | 1 | 14% |
| efficiency | week 4 | 18 | 12% | 9 | 50% | 9 | 50% |
| | week 11 | 12 | 10% | 8 | 67% | 4 | 33% |

The *participation* theme captures whether respondents felt that Gameface encourages or discourages active participation in class. It was found in ~40% of responses at both timepoints, and sentiment was almost universally positive. Students felt that Gameface heavily incentivised active participation, and encouraged the whole class to participate rather than relying on a few volunteers. However some respondents felt that assessment made participation feel forced.

> *"Makes me engage in the class, normally I would never offer to answer a question or be so engaged without the opportunity to gain marks." - week 4*

> *"[Gameface] encourages participation and shares participation between all students." - week 4*

> *"It would increase interaction in classes and actually make students (including myself) think and participate." - week 4*

> *"It threatens marks to force class participation creating an anxious environment. It's not something I look forward to." - week 4*

> *"It did not feel conducive to organically asking for a volunteer or asking questions as we all just wanted to get the marks and go home." - week 11*

The *learning* theme captures whether Gameface was perceived as having pedagogical value. It occurred in 25% of responses in week 4 and 33% in week 11. Sentiment was 84% positive in week 4 and 92% positive in week 11. Many students felt that the tool helped them stay focused and active during the class (i.e. active learning). Others highlighted that it strongly encouraged them to prepare well before coming to class (i.e. flipped learning). Students also felt that the tool helped both students and tutors know which areas were mastered well or required further attention (i.e. feedback). Most of the negative comments focused on individual tutors who did not take the time to address knowledge gaps that were uncovered during quizzes. Another concern was that students may focus more on assessment and scoring points than taking advantage of learning opportunities, but this concern was only present in the Week 4 survey.

> *"It generated good class involvement and helps students stay focused and switched on for the content that was covered that week" – week 4*

> *"It really forces me to stay on top of the lecture material every week and not fall behind - week 4*

> *"It is a tool in which the tutor can see what areas students are struggling with and it also allows students to compete against each other and try their best to earn points which helps them to learn from their mistakes." - week 4*

> *"Also its really IMPORTANT that teachers explain why the result is NOT the wrong answer if a lot of students get it wrong. [Tutor] does do this well, but the other teacher I've seen doesn't bother to explain why we got it wrong and so we are continually confused lost and stressed." - week 11*

> *"I find that it is a great learning tool however with the pressure of it being assessed I personally am only focused on getting answers right (as to ensure I receive the best mark possible) rather than taking advantage of all the learning opportunities." - week 4*

> *"I need the time in tutorials to understand what I don't know. I am prepared to make mistakes or show up to class with issues or questions without worrying about how it will affect my final mark." - week 4*

The *enjoyment* theme captures whether respondents felt that Gameface was fun and enjoyable. It was mentioned in 13% of responses in both week 4 and week 11, and most were positive. Many respondents praised it for making labs feel more playful. A minority felt that assessment created anxiety and stifled experimentation.

> *"It is interactive and fun to use." - week 11*

> *"Making labs interactive and game-like makes it more engaging and fun for students." - week 4*

> *"I didn't really enjoy using Gameface as it added a level of extra pressure in labs, especially in weeks where I felt like I didn't fully understand the concept. I would have rather spent that time getting help to figure things out." - week 11*

> *"Forced contribution but also simultaneously becomes an added pressure in our already pressurised environment." - week 11*

The *collaboration* theme captures whether Gameface encouraged or discouraged students from working with each other. It was mentioned in 14% of responses in week 4 and 9% in week 11, with 90% positive sentiment at both timepoints. The positive responses show that the tool provided opportunities for students to meet with and learn from each other, and the automated matchmaking eased anxiety about initiating connections. It also helped students more confidently

contribute and be visible in the classroom. Negative responses focused on a lack of agency and continuity regarding who students were paired with.

> "Gameface is fun and is a good way for the class to participate even if you are a shy person. I like that it puts you in pairs with a random person each week so that you can learn to talk and meet new people who might have better answers than what your existing friends can help you with." - week 4

> "It also helps develop communication skills and confidence with speaking in groups/answering questions in front of the class." - week 4

> "Helps ease the anxiety of needing to choose a partner for labs/an activity when you can have a system that automates it for you, and the filtered groups helps provide the opportunity for a person with a lower performance in the subject to be in a group with people with a higher performance so that they can hopefully learn better and improve their own marks." - week 4

> "It also introduces a low stakes form of pressure for the student to get more comfortable adding to class discussions, answering questions etc which will benefit them in their degree and in work life." - week 11

> "It makes you more antisocial because you're on your phone, harder to make good friends because you're always switching around partners." - week 11

The *fairness* theme captures whether Gameface was perceived as a reliable basis for formal assessment. The prevalence and sentiment of this theme varied greatly between the two surveys. In week 4 it occurred in 15% of responses, of which 74% were negative. Much of the negative sentiment could be attributed to early teething issues that required scores to be recalculated, but more fundamental concerns focused on transparency and the need to account for diversity. Some students felt the scores were opaque and difficult to understand. Others were concerned that chosen volunteers were given an unfair advantage, and this form of participation unfairly penalised those with social anxiety. These concerns appear to have been greatly alleviated by the updates made to Gameface during the session (i.e. the introduction of health and armour, the provision of detailed breakdowns, and the changes made to how volunteers were selected), given that prevalence of this theme dropped to 6% in week 11, and sentiment towards it was 86% positive.

> "It is not clear at all how points are allocated and as this has an effect on the overall mark for students it is stressful when bugs and such are encountered and change our grades." - week 4

> "Only a few get a chance in the class to say things or to volunteer whereas others don't get any and hence their marks are less than the others." - week 4

> "Gameface is very harsh with its marking making you lose significant % for small mistakes and creates difficulty for people with social anxiety to be able to get high marks as not everyone wants to have their grade determined on their ability to raise their hand in class." - week 11

> "As it is a more immersive way of learning, where participation is key instead of just focusing on the results. To know at least a student tried even if they did not 'succeed'" - week 4

The *efficiency* theme captures whether Gameface made the class run more or less smoothly. It occurred in 10-12% of responses. Sentiment went from 50% positive in week 4 to 66% positive in week 11. On the plus side it was intuitive and made quizzes and group formation smoother. On the negative side it introduced friction due to the need to switch devices and displays frequently.

> "The Gameface UI has been very intuitive so far." - week 4

> "Gameface enhances learning with engaging quizzes, fosters collaboration through group formation and saves time." - week 11

> "I don't like the multiple interfaces we are using at the same time. It takes time to change from one to another. In order to have all of them at the same time, I have to use multiple devices. It makes me feel less concentrated and sometimes a bit confused. Also the tutor has to constantly change from Gameface to others, making things a bit complicated. For example, asking for a volunteer takes multiple steps to complete on Gameface." - week 4

# Discussion and conclusions

This study provides an example of how to apply gamification to achieve fine-grained measurement and assessment of classroom engagement without losing student endorsement. Our thematic analysis shows that Gameface was perceived by students as improving behavioural engagement (e.g. encouraging students to prepare well for class and be active participants during class), cognitive engagement (e.g. helping students stay switched on and focused), emotional engagement (e.g. maintaining a playfully competitive tone) and social engagement (helping students work with and learn from their peers). The level of endorsement received by Gameface was particularly high when one considers that its use as an assessment item effectively forced attendance in a post-covid era in which teachers (including the author) are finding it difficult to encourage students to return to campus (Detoni et al., 2024).

The study is not without limitations, the largest of which is that more robust measurement is needed before we can confidently say that Gameface successfully promotes and assesses engagement without undermining it. The endorsement and feedback it received from students is promising, and the participation and understanding scores provide insight into student's behavioural and cognitive engagement. However our only evidence regarding emotional and social engagement are isolated comments in which students describe the tool as fun and healthily competitive, but also somewhat stressful and inhibiting. The design changes introduced during the study lessened participant's concerns, but did not eliminate them entirely. Additionally, the study focuses exclusively on a single subject, so generalisability may be questioned. We are currently planning a larger study which would employ validated scales of behavioural, cognitive, emotional and social engagement and recruit students/subjects from multiple disciplines. Teachers and coordinators who are interested in participating in such a study (by deploying Gameface within their classrooms) are encouraged to contact the author.

# References

Alonso-Tapia, J., Merino-Tejedor, E. & Huertas, J.A. (2023). Academic engagement: assessment, conditions, and effects—a study in higher education from the perspective of the person-situation interaction. *Eur J Psychol Educ,* 38, 631–655.

Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology in higher education: A systematic evidence map. *Int J Educ Technol High Educ*, *17*, 1-30.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qual Res Psychol, 3*(2), 77-101.

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innov Educ Teach Intl*, *61*(2), 228-239.

Detoni, M., Allan, A., Connelly, S., Summerfield, T., Townsend, S., & Shephard, K. (2024). University teachers' perspectives on student attendance: a challenge to the identity of university teachers before, during and after Covid-19. *Educ Res Pol Pract*, 1-19.

Gibbs, G. (2006). Why assessment is changing. In *Innovative Assessment in Higher Education*. Routledge.

Leber, J., Renkl, A., Nückles, M., & Wäschle, K. (2017). When the type of assessment counteracts teaching for understanding. *Learn Res Pract*, *4*(2), 161–179.

Mandernach, B. J. (2015). Assessment of student engagement in higher education: A synthesis of literature and assessment tools. *Intl J Learn Teach Educ Res*, *12*(2), 1-14.

Marcum, J. W. (2000). Out with motivation, in with engagement. *National Productivity Review*, 18, 57-59.

Wang, M. T., Fredricks, J. A., Ye, F., Hofkens, T. L., & Linn, J. S. (2016). The math and science engagement scales: Scale development, validation, and psychometric properties. *Learn Instr*, *43*, 16-26.

## Copyright statement