TASO Transforming Access and Student Outcomes in Higher Education

Final implementation and process evaluation report: University of Huddersfield – Score As I Learn (SAIL)

September 2024

Authors:
Dr Jarosław Bryk, Strategic Teaching and Learning Team, University of Huddersfield
Steve Bentley, Strategic Teaching and Learning Team, University of Huddersfield
Dr Jane Wormald, Strategic Teaching and Learning Team, University of Huddersfield
Dr Ruth Stoker, Strategic Teaching and Learning Team, University of Huddersfield

QA:
Christoph Koertbitz, TASO

Table of contents

1. Executive summary

1.1. Project Team

The Score As I Learn (SAIL) initiative was first proposed by the Head of Department of Engineering and Technology (CEET) in the School of Computing and Engineering (SCE) at the University of Huddersfield, in 2019, and its implementation and department-wide rollout was subsequently modified by the academic community in the SCE. The team that prepared and analysed the required datasets for this IPE report included members of the Strategic Teaching and Learning Team (STLT), SCE and university staff, in particular:

- Dr Jaroslaw Bryk, University Teaching Fellow and Data Analyst, STLT
- Steve Bentley, Learning Technology Advisor. STLT
- Dr Keith McCabe, Head of Planning and Business Intelligence
- Dr Jane Wormald, Pedagogic Lead, STLT
- Dr Ruth Stoker, Director of Strategic Teaching and Learning
- Matt Mills, Director of Student Services
- Prof. Jane Owen-Lynch, Pro-Vice Chancellor for Teaching and Learning
- Prof. Artur Jaworski, SCE
- Prof. Andrew Crampton, Associate Dean for Teaching and Learning, SCE

1.2. Project description

1.2.1. Intervention being evaluated

SAIL is a series of weekly, week-long assignments offered to all students in all modules in the Department of Engineering (CEET), SCE. The assignments are voluntary and summative: 11 weekly assignments contribute 3% each towards the module mark, but only 8 best results are included, for a maximum of 24% of module overall mark derived from SAIL assignments. The assignments are delivered and marked automatically *via* the virtual learning environment (VLE) platform and are designed to cover foundational knowledge and topics being taught in a given week. The principal motivation behind the introduction of SAIL was to increase students' engagement with the taught material and develop their academic study habits.

1.2.1.1. Target group

The SAIL initiative in 2021-22 and 2022-23 included all undergraduate students in CEET.

1.2.1.2. Number of students involved

There were approximately 600 students included in the analysis in 2021-22 and 500 students in 2022-23, with approximately 1000 students each year in a control group, from the Department of Computer Science (CEI) in the same School.

1.2.1.3. Implementation

The implementation of SAIL involved creating a VLE -based repository of questions by tutors, targeting the learning in a given week for every module, such that a set of 5 questions could be drawn randomly for each student in every week and module. Utilising the functionality of the VLE, each assignment remained open for a week and allowed each student one attempt at answers. Following submission by each student, the questions were then marked automatically by the VLE and the results released to students within 24 hours after the weekly deadline. Therefore, after the initial effort of setting up the questions, the initiative is implemented universally and automatically, with maintenance and update tasks required from staff.

1.2.2. Brief description of the IPE

This Implementation and Process Evaluation (IPE) forms part of the TASO Institutional Data Use project and complements an Impact Evaluation (IE) conducted by an independent team from Staffordshire University. This evaluation provides context for the IE report, which comprises a quantitative analysis of the program's effectiveness in achieving its objectives. The IPE focuses on:

- Assessing the operational aspects of the SAIL initiative
- Examining the implementation of the programme
- Identifying the factors influencing the implementation processes
- Discussing our ability to evaluate the effectiveness of the programme.

Principally, this IPE analyses how uniform and universal the implementation of SAIL was in terms of student participation and the characteristics of the assignments. In particular, it asks whether any of SAIL's features were potentially disadvantaging any particular group of students. Finally, given the quantitative IE outcome of no significant differences in attendance between the SAIL and non-SAIL groups, it discusses high-level limitations of its implementation as well as challenges of evaluating student engagement.

1.2.3. Key findings

The principal findings of this IPE are:

- Confirmation of widespread availability of the SAIL and high completion rates of the quizzes.
- Predominantly multiple-choice nature of quiz questions, with over half of quizzes including more than one type of questions.
- An inconsistent implementation of quizzes' availability: while all modules had quizzes available throughout the year, only half of the quizzes were available for the expected 7 days and/or included questions randomly drawn from a pool.
- Considerable variation in average SAIL scores between different modules but little difference in SAIL scores week-by-week.
- Lower SAIL scores for students in the low participation group compared to the high participation group.
- Overrepresentation of students with BTEC-like entry qualifications and students from areas with low scores on the Indices of Multiple Deprivation (IMD) among the low participation group compared to the overall population of SAIL participants. IMD is commonly used as a proxy measure of socioeconomic status particularly, those in the bottom two quintiles, which represent the most deprived areas.
- No differences in distribution of ethnicities (White vs. Black and Minority Ethnic [BAME]) between low and high participation groups.

1.2.4. Key conclusions

The IPE demonstrated that student engagement with SAIL was high, with a majority of students clearly engaging with SAIL week to week, with relatively consistent scores throughout the term. However, there was uneven implementation across the school due to some technical issues related to VLE functionality. We also found evidence that students with BTEC-like entry qualifications and students from lower socioeconomic background tend to be overrepresented among the students who completed between 1-8 quizzes compared to overall SAIL participants and that these students also achieve lower SAIL scores compared to students from the high participation group. These findings suggest that the SAIL approach on its own may not be sufficient enough to support all students in developing their academic study skills to bridge the IMD and/or entry qualifications gap at Huddersfield.

However, the key conclusion from the analysis is that there is insufficient quantitative and lack of any qualitative data for an in-depth evaluation of the motivations and mechanisms driving students' and staff's participation and engagement with SAIL.

2. Introduction

2.1. Background and rationale

The School of Computing and Engineering (SCE) at the University of Huddersfield has approximately 1500 undergraduate students across all years of study. The school comprises two departments, Department of Engineering and Technology (CEET) and Department of Computer Science (CEI), with a yearly intake of approximately 230 and 300 students, respectively. Both Departments have a similar demographic of (UK) students, with roughly 60% of students being of White ethnicity and the largest minority students, of British-Pakistani ethnicity, making a further 23% of students. Over 50% of students in each department come from lower socioeconomic background (IMD quintiles 1 and 2).

There are some differences in qualifications on entry of students in the two departments. In CEET, students with triple A Levels make up 28% of the population, followed closely by triple BTEC qualification at 22%. In CEI, BTEC qualifications dominate at 37%, with A Levels making 22%. The mixed Level 3 qualifications are at 30% in CEET and 26% CEI. Similarly for the assessment structure: coursework is the most common assessment type in the two departments (47% in CEET, 37% in CEI), followed by tests/exams at 32% in CEET and 24% in CEI). Portfolios are markedly higher in CEI over CEET (24% vs 5%). These characteristics of the schools have been relatively stable in the last four years, but the numbers above are provided for the academic year 2021-22.

Despite these similarities, the two departments differ in terms of outcomes, with the attainment gap (difference in the proportion of 2:1s and 1st between White and BAME students) in CEET being double of the CEI (13 percentage points vs 6 percentage points) though both gaps are lower than sector average (17 percentage points).

Our previous work demonstrated that students with no-A Level qualifications on entry were most strongly associated with student outcomes in the SCE, with BTEC and mixed Level 3 qualifications particularly strongly associated with lower module marks (7.9 and 9.7 percentage points lower, respectively, compared to triple A Level, $p < 0.001$).

Gender and ethnicity had no significant association with outcomes, however students in IMD quintile 1 had a significant negative association with module marks (3.3 percentage points lower, $p < 0.05$). These patterns were also clearly visible when comparing distributions of module marks across demographic characteristics and other factors, such as assessment type, where we observed exams being associated with lower marks for BTEC, IMD quintile 1 and ethnic minority students in CEET but not in CEI.

Supported by other evidence, including earlier projects on transitioning and belonging (such as “Flying Start”), distributions of module marks from classes and assessments

delivered during the lockdowns period, contact with six-form teachers and interviews with students, we hypothesised that, on average, students with non-A Level qualifications require additional support to engage with their courses and academic practice expected at Huddersfield compared to their peers with A Level qualifications.

Low-stakes summative assessments, dubbed SAIL (Score As I Learn), were originally introduced in the CEET in September 2019 as a means of improving attendance and engagement with class contents by international students. The necessity of switching to asynchronous delivery due to the onset of the COVID19 pandemic in early 2020 provided a stimulus to expand the SAIL approach to all modules in the CEET from September 2020, with a view of boosting students engagement with the (then online-only) course content. This switch surreptitiously also reduced the contribution of the exam component on modules, a factor that our post-lockdown analysis of module marks demonstrated had a significant positive effect on minority students' marks compared to "standard" delivery, which was dominated by one-off, high-stakes assessments carrying majority of mark weight for a module.

2.2. Intervention aims and objectives

SAIL was introduced from September 2020 as part of the institution's COVID-19 emergency procedures to fulfil three main functions:

(i) to guide students' learning via the setting of week-long assignments (usually Virtual Learning Environment (VLE) quizzes), testing their understanding of the material of which a large proportion was delivered online;

(ii) to provide a measure for module teams of the actual student engagement via VLE reports on the uptake of SAIL quizzes and marks;

(iii) to help module teams identify parts of the material that may not have been commanded sufficiently well by the students.

2.3. Detailed description of the intervention (to enable replication)

At its heart, SAIL approach is a series of week-long summative assignments, delivered through the VLE (in the form of quizzes), where each assignment carries a small fraction of overall module mark (3-5%). The assignments are released at the start of the week and are open for a week, with each student receiving a randomised selection of questions related to that week's classes. There is no time limit on the assignment once it is started other than the overall weekly deadline, but each student has only one opportunity to take the quiz, with no extensions possible. The results are released to students automatically within 24 hours after the deadline in each week.

In its first full year of SAIL operation (2020/21), each student was given 10 weekly assignments weighted at 5% of the module mark per assignment, but only the best six counted towards the final module mark, for a total of 30%. In the second year (2021-22), these proportions changed to 8 out of 11 assignments, each at 3% of the module mark, for a total of 24%. This approach was motivated by the need to accommodate extenuating circumstances that could prevent students from submitting any given assignment, particularly given the lack of option to extend any deadline.

The questions are explicitly designed to target foundational knowledge delivered in a given week, with the purpose of encouraging engagement with the material and students checking their own level of knowledge and understanding of the material. Simultaneously, the level and quality of student engagement with the SAILs (how many students submit them and their results) provide a check for the module leader of students' command of the material and a proxy of their use of class and related materials throughout each week. Potential outliers in terms of results or level of SAIL uptake each week may prompt a module leader to run a feedback or revision session, already in the following week after the challenging material was first introduced. Additionally, SAILs are intended to be habit-forming for students to develop their academic practice and support the principle of spaced repetition, aiding their learning.

Finally, SAILs, after the initial work of setting up of weekly question pools (which can be as simple or as imaginative as a platform allows), are relatively low maintenance and rely entirely on automatic mechanisms implemented in the VLE, which in turn facilitates extending and tweaking of questions and answers, as necessitated by taught materials.

2.4. Brief description of the evaluation (including the IE) and what this report covers.

As specified in our Theory of Change, the major aim of SAIL was to increase student engagement, and ultimately course and module outcomes. For the purpose of the IE, we defined engagement as any parameter informing of student participation in the academic process: module attendance, submission of assignment and their timing (e.g. early/late submission), use of the VLE (e.g. time spent in VLE, number of sessions), as well as related measures such as use of lecture capture (e.g. time spent watching) or library facilities (e.g. number of visits). The IE uses several measures of students' outcomes, such as module and assessment marks, degree outcome, continuation and progression, and investigates the effect of various demographic characteristics, such as ethnicity, socioeconomic status, qualifications on entry, among others, on these outcomes.

With mixed effects modelling, we aimed to uncover the effectiveness of SAIL approach on various measures of outcomes and also point to factors modulating these effects in various sub-populations of students.

In this report, we detail the implementation of the process of running SAIL in CEET. However, due to limitations of the data and its quantitative nature, we concentrated on the student side of the implementation.

3. Methodology

3.1. Statement of research questions.

The main research questions this IPE attempts to answer are:

- IPE-RQ1 How was the SAIL initiative implemented?
- IPE-RQ2 How did the students engage with SAIL initiative?

3.2. Overview of IPE design and methods.

For the IPE, we used the VLE-derived data on the number of SAIL assessments submitted per module and student, and cross-checked it with student demographic characteristics using a subgroup analytical approach.

3.2.1. Specific research questions related to IPE

IPE-RQ1 How was the SAIL initiative implemented?

a) Have SAIL quizzes been run across all modules in CEET?
b) Have SAIL quizzes been run for all 11 weeks in all modules?
c) Have SAIL quizzes' availability in the VLE been timed correctly?
d) Do all SAIL quizzes have similar content?
e) Have SAIL quizzes scores been distributed similarly across modules?
f) What were the main issues with the implementation?

IPE-RQ2 How did the students engage with SAIL?

a) Have all students participated in SAIL?
b) Do all student demographic groups engage with the SAIL similarly?
c) Are there differences in levels of participation in SAIL among different student demographic groups?
d) What are the patterns of engagement for these demographic groups?

3.2.2. Research methods

As the SAIL was implemented entirely with the VLE, the delivery of the implementation and students' adherence to its principles was checked using VLE records accessed through the DOMO back-end.

This data was then matched to student records available in the Planning office to extract various characteristics of students such as ethnicity, qualifications on entry and socioeconomic status (as measured via Index of Multiple Deprivations based on home postcode). These subgroups were compared for their participation level in SAIL and the outcomes of SAIL that they had achieved, to identify whether some groups of students are over- or underrepresented based on their degree of participation in SAIL.

3.2.3. Sample, data sources (size, demographics, recruitment, data collection)

For the IPE we have used the same population of students that was part of the IE, approximately 1100 SAIL participants across two academic years, 2021-22 and 2022-23. The demographic characteristics, however, are only available for home (UK) students.

3.2.4. Details of dosage, compliance, fidelity, and usual practice.

Because SAIL was introduced obligatorily across all modules in CEET in a uniform fashion of 11 weekly quizzes, the variation in the adherence to its implementation could practically only vary due to technical error in setting up the quizzes between different members of staff or errors of the VLE itself. Following a period of setting up the quizzes, their delivery and marking is automatically handled by the VLE and thus universal.

Since there is no formal requirement for any student to take any part in the SAIL, there may be differences in compliance between students. However, the significant mark component of the assignments (24%) and very low barrier to entry are good motivators for participation.

Given that the format and timing of SAIL was unchanged between 2021-22 and 2022-23, investigation of the effects of the adaptations of the intervention is not possible. Similarly, as the SAIL is now embedded in usual practice across the department, in an automated fashion *via* the VLE, the evaluation of the usual practice does not apply.

3.3. Ethics

The project was submitted for approval to the School of Applied Sciences Research Integrity and Ethics Committee, a committee local to the project's Principal Investigator, according to the standard University of Huddersfield policy, on 6 October 2023, and received approval on 19 December 2023 under reference number SAS-SRIEC-19.12.23-2.

4. Findings

4.1. Summary of the findings

IPE-RQ1 How was the SAIL initiative implemented?

4.1.1. Timing, content and availability of quizzes

The implementation question of SAIL primarily concerns the number and timings of all the SAIL quizzes. Our analysis of VLE records indicates that the delivery of the planned quizzes was good, albeit not perfect.

In 2021-22, 90% (84/93) of modules had 11 SAIL quizzes set up, and in 2022-23 this proportion was 98% (95/97). In both years, for a small number of these quizzes (19/190, 10% altogether), we find evidence of imperfect setup, e.g. duplication of quizzes, yet those quizzes were still delivered correctly. In six additional modules the quizzes were set up incorrectly, with either too small (9-10) or too large number of quizzes (>12). But overall, in our view this is a very good result, indicating a near-universal implementation of SAIL. A few modules without quizzes were those for which they were not applicable, such as project modules.

Quizzes appear to be set up with a reasonable variety of questions. According to VLE's classification of the question types, 75% of all questions are of multiple choice type, but over 53% of all quizzes mix more than one type of questions:

Type of question	Proportion
Multiple Choice Question	74.8%
True or False Question	13.5%
Arithmetic Question	4.7%
Multi-Select Question	4.6%
Fill in the Blanks Question	1.1%
Short Answer Question	0.4%
Matching Question	0.4%
Significant Figures Question	0.2%
Ordering Question	0.1%
Long Answer Question	0.1%
Multi-Short Answer Question	0.1%

Another way of looking at the consistency of quizzes' questions is to compare SAIL scores across different modules, as similarly difficult quizzes in different modules should result in similar distribution of scores. But when we look at average SAIL scores per module, we see a relatively large variation each year, with the highest-scoring module roughly double of the lowest scoring one (Fig. 1).

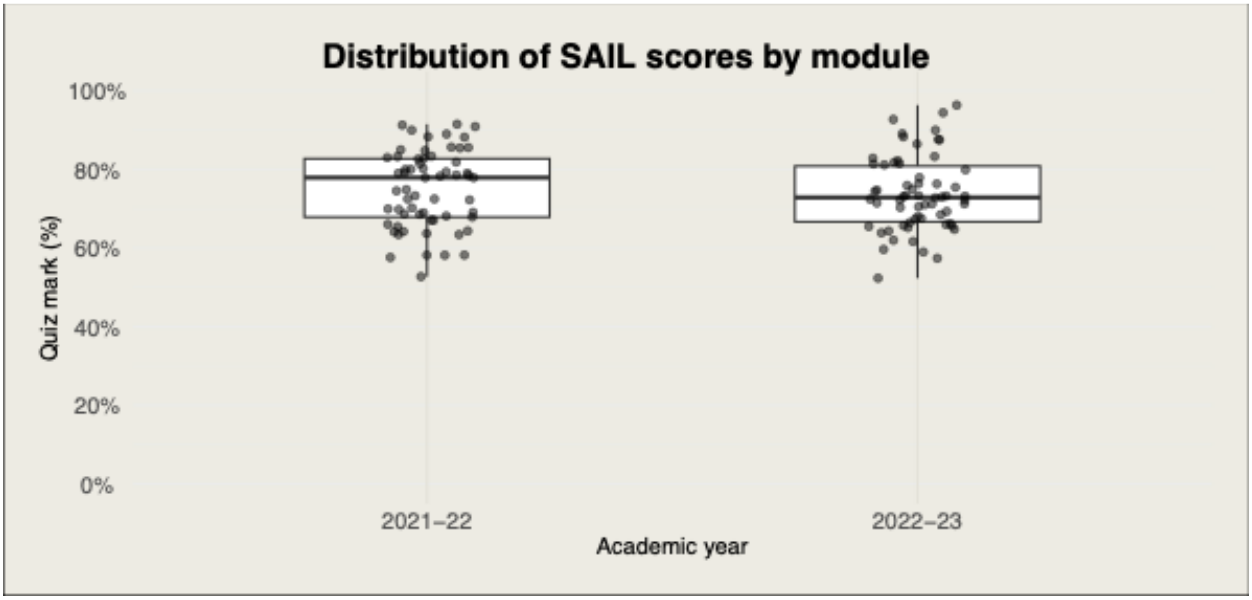

Figure 1. Box and whiskers plot for the distribution of average SAIL scores for each module, for all quizzes and students. Only modules that had 11 quizzes set up are included.

It is difficult to interpret this observation based on this data alone. As SAIL was meant to be based on questions requiring only recognition or remembering information rather than analytical and evaluative skills (see the SAIL theory of change), it could be expected that most quizzes would have a similar level of difficulty and therefore their scores would be similar. On the other hand, it is possible that various topics have higher perceived or actual level of difficulty for the students, even at the foundational level of questions, and the scores above reflect this variation. Answering this question would require consultations with subject experts and module leaders who set the questions for each quiz.

In terms of timing of each quiz, the pattern differed from what we expected. Instead of each quiz starting and ending on the same weekday for a 7-day availability pattern, we find that quizzes do not have a consistent start and end date. We suspect that each quiz was made available following a class in a given week, and therefore – because of the variable weekly timetable – the quizzes were available on different schedules.

Only about 53% of quizzes (294/553) were available for 7 days in 2021-22, and 70% (678/975) in 2022-23. Majority of the rest (260/1528, 17% altogether) was opened for a longer period of up to three weeks. There was also a long tail of a number of quizzes each opened for a much longer period of time, consistent with the quizzes being incorrectly “rolled over” from the previous year, and a number of quizzes with neither start or end date specified (108/1528, 7% altogether). Similarly, approximately half of the quizzes did not appear to randomly draw questions from a pool, likely a consequence of incorrect setting up the pools. Finally, 87% and 88% quizzes in 2021-22 and 2022-23, respectively, have correctly allowed only one attempt in the VLE.

This indicates various inconsistencies in setting up the quizzes and raises questions whether this may affect who and when completes the quizzes during the term. We suspect these inconsistencies are a result of individual staff errors or omissions in setting up the quizzes. This is relatively easy to overcome, as the entire process: setting up the start, end date, attempts allowed and duration of each quiz can be automated and overseen through a VLE, *if* sufficient development and maintenance resources would be available.

IPE-RQ2 How did the students engage with SAIL?

4.1.2. General patterns of student engagement with SAIL

On the students’ side of SAIL implementation, we have checked the proportion of students who submit the quizzes, for each number of quizzes submitted. To standardise the number of students per module (and therefore number of possible quiz submissions

per module and the denominator in our calculations), we used the number of students in each module who had submitted any assignment apart from SAIL in the term.

On average, in the modules with exactly 11 quizzes available, each student submitted 10 quizzes across all their modules (this number is the same in both years), and the completion rate (proportion of students who submitted a quiz) in each week had a pattern illustrated on Fig. 2.

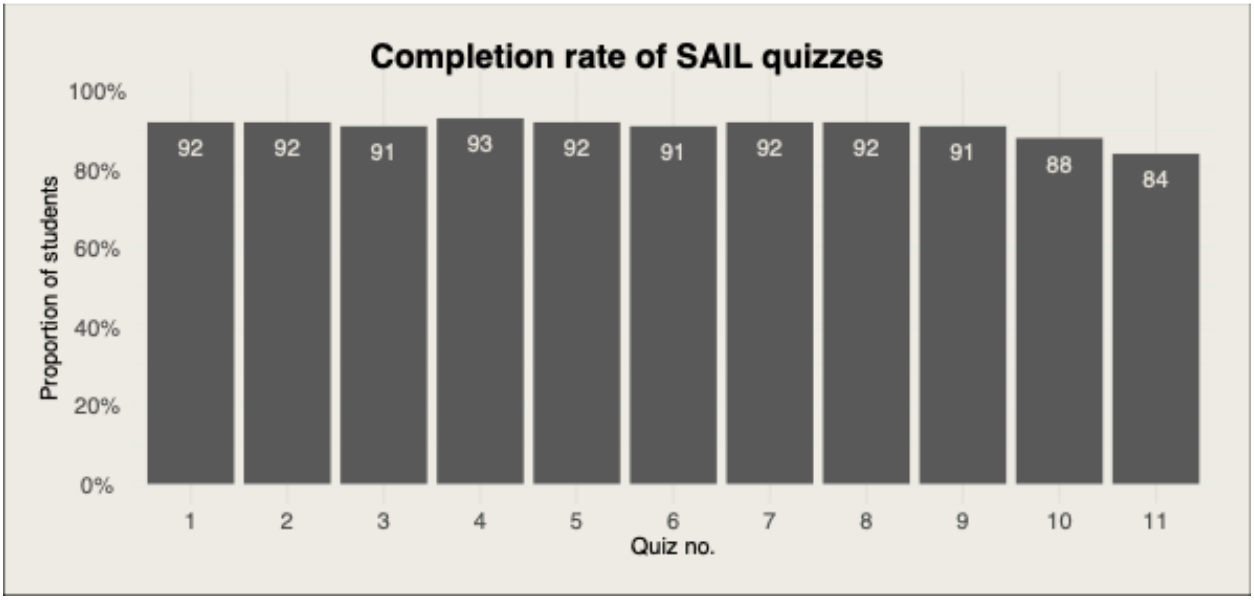

Figure 2. SAIL completion rate: proportion of students who submitted a quiz in a given week. Data aggregated across both academic years (2021-22 and 2022-23).

All but the last two weeks of SAIL have over 90% completion rate, which then drops down below 90%. It still is high in week 11 at 84%, but we interpret it as suggestive of some strategic planning by students, who appear to submit the minimum number of quizzes required for full marks, plus one or two, and then disengage from the process towards the end of the term.

Because there is not a common maximum number of SAIL quizzes for all students, as they may take different numbers of modules, not all of which have SAIL, it is difficult to establish the distribution of student submissions of SAIL quizzes. To approach this problem, we calculated an average number of quizzes each student submitted across all their modules and plotted the resulting distribution in Fig. 3.

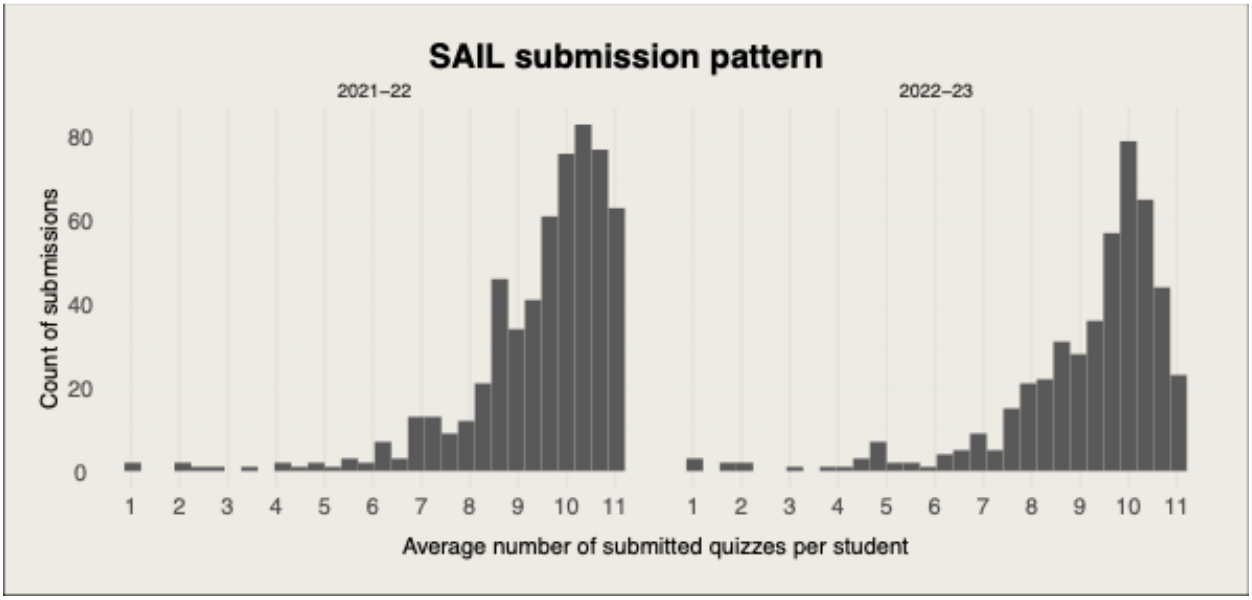

Figure 3. SAIL submission pattern: distribution of the average number of submitted quizzes per student.

Overall, the level of student participation in SAIL was impressive, likely due to the nature of the process, with low barriers to entry and offering large academic incentives. More generally and more importantly, it is an indicator that students' outside-of-classroom effort *can* be harnessed repeatedly throughout the term, which itself bodes positively for other educational interventions.

There was some, but surprisingly little, variation in student scores from each SAIL when aggregated across all modules. At the same time, there was a wide distribution of marks across all quizzes, indicating very low scores for some students, each week (Fig. 4).

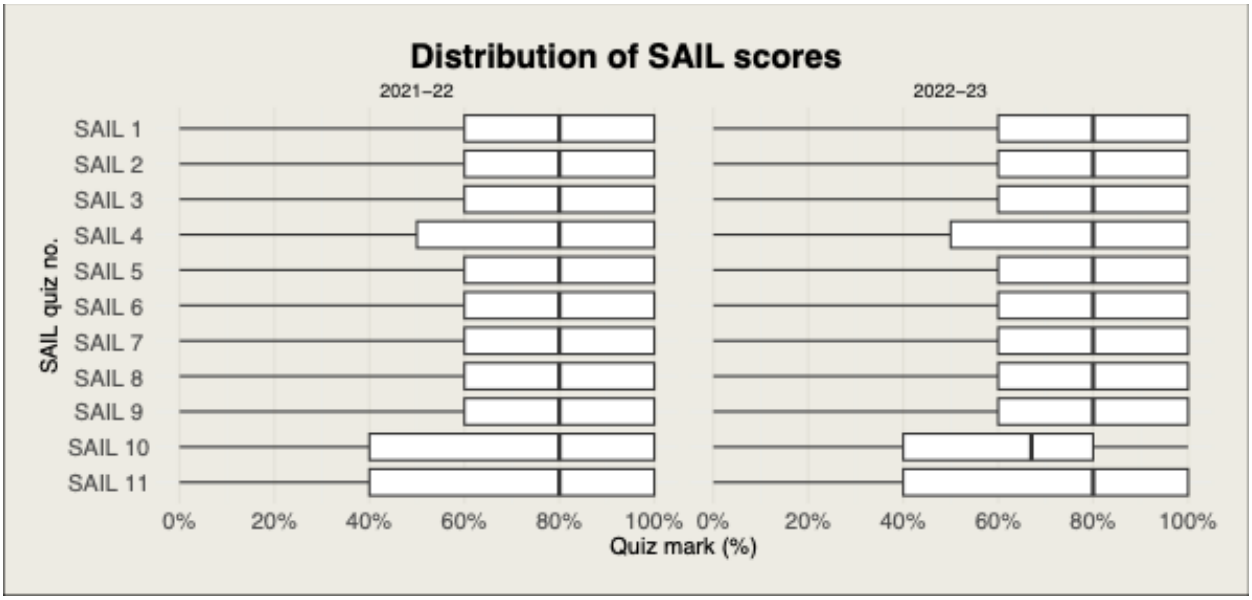

Figure 4. Distribution of SAIL scores for each quiz for all students and across all modules.

4.1.3. Patterns of engagement with SAIL between different student subpopulations

We can finally ask whether, despite the uniformity of scores and excellent participation rate in SAIL, there are any differences in which groups of students engage with the process. Based on the distribution in Fig. 3, we created two groups of students, with “low participation” – who submitted between 1-8 quizzes, and “high participation” – who submitted 9-11 quizzes, and asked whether the two groups achieved different SAIL scores. The data suggests that this is indeed the case, with low participation group achieving significantly lower scores that high-participation group (Fig. 5).

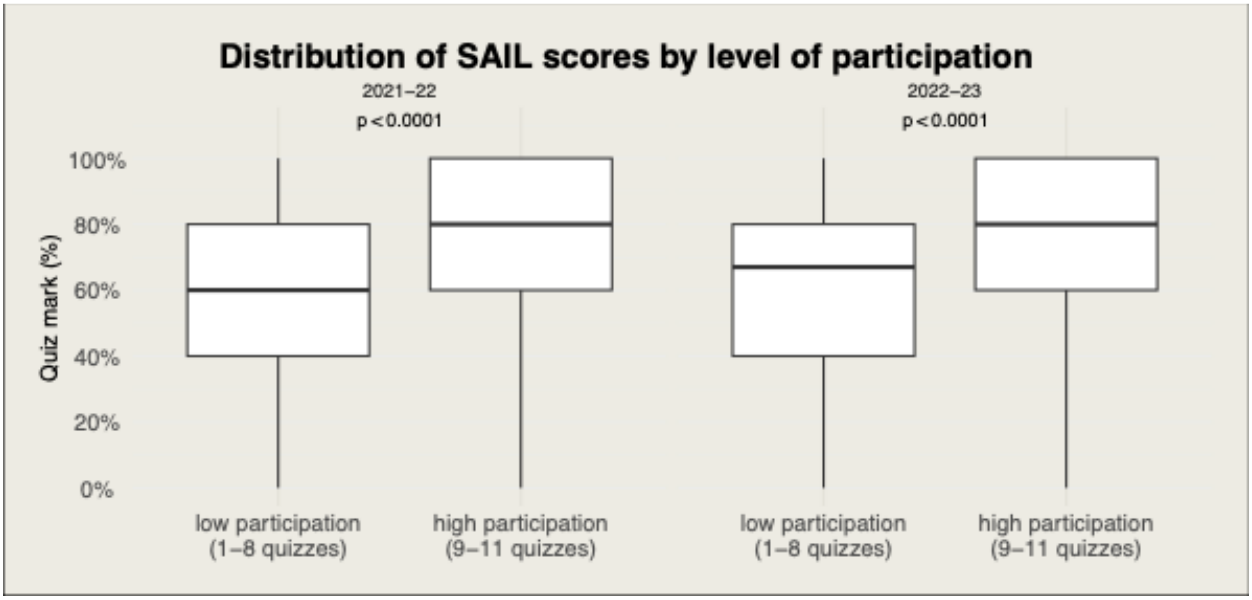

Figure 5. Distribution of SAIL scores by level of participation. Low participation indicates students who completed 1-8 quizzes, high participation indicates students who completed 9-11 quizzes. P values were obtained by running a T-test.

Below, three pairs of plots illustrate the baseline (proportions of SAIL students by their ethnicity, entry qualifications and index of multiple deprivation levels (a proxy for socioeconomic status)), paired with how they are arranged between the low and high participation categories (Fig. 6-8).

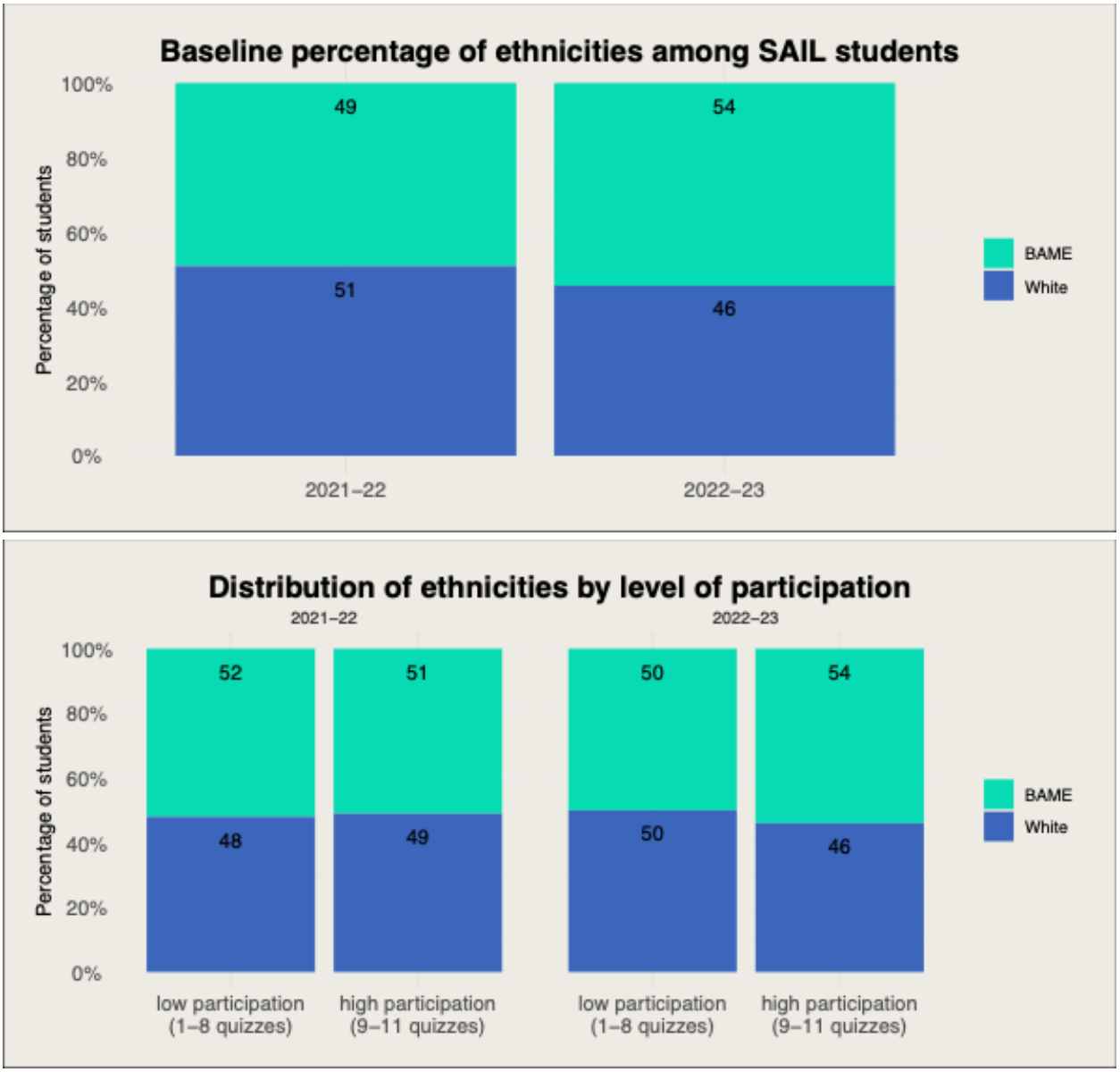

Figure 6 Percentage of White vs. BAME students among all SAIL participants (top panel) and split by the level of participation in SAIL (bottom panel).

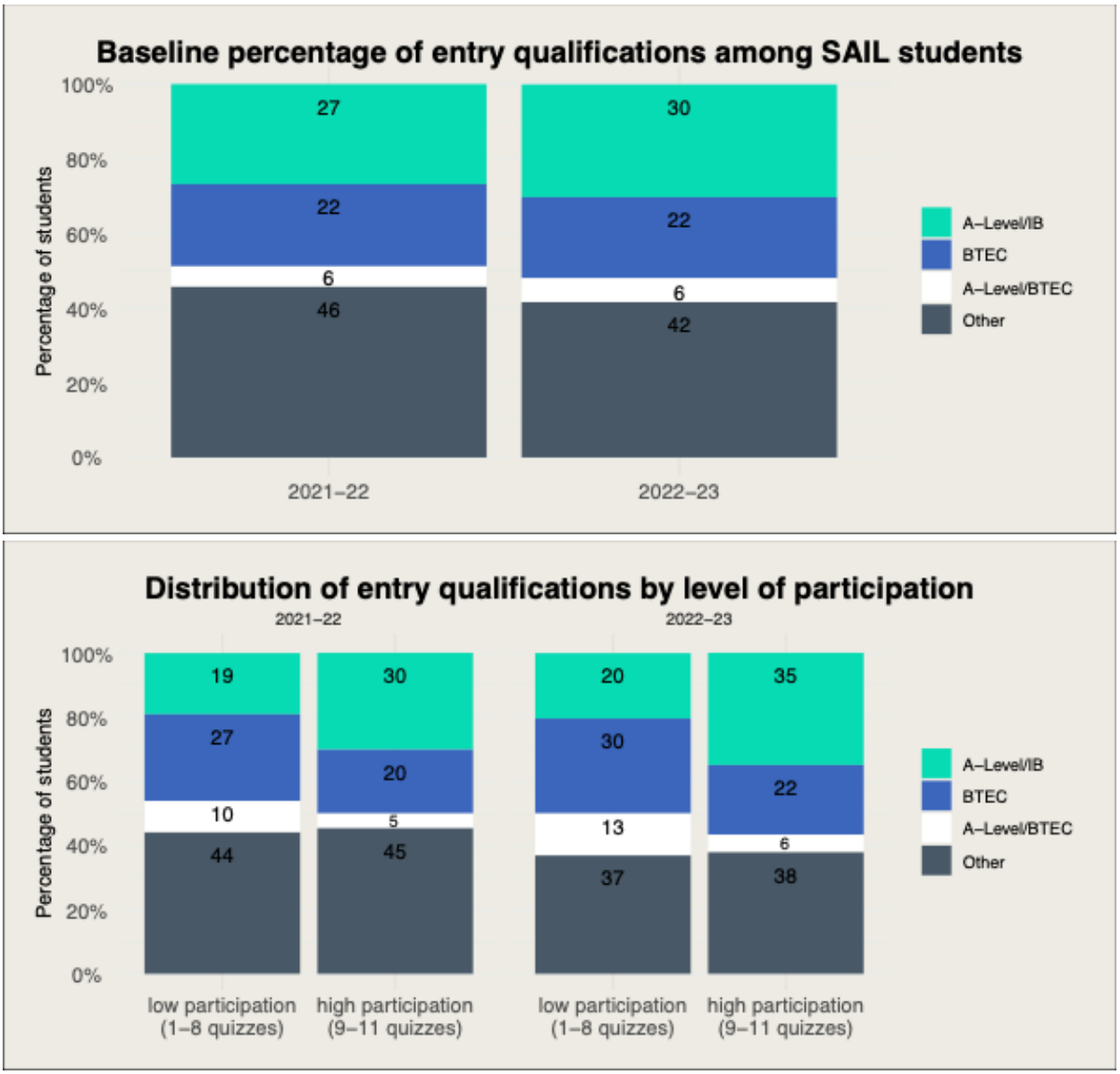

Figure 7 Percentage of entry qualifications among all SAIL participants (top panel) and split by the level of participation in SAIL (bottom panel).

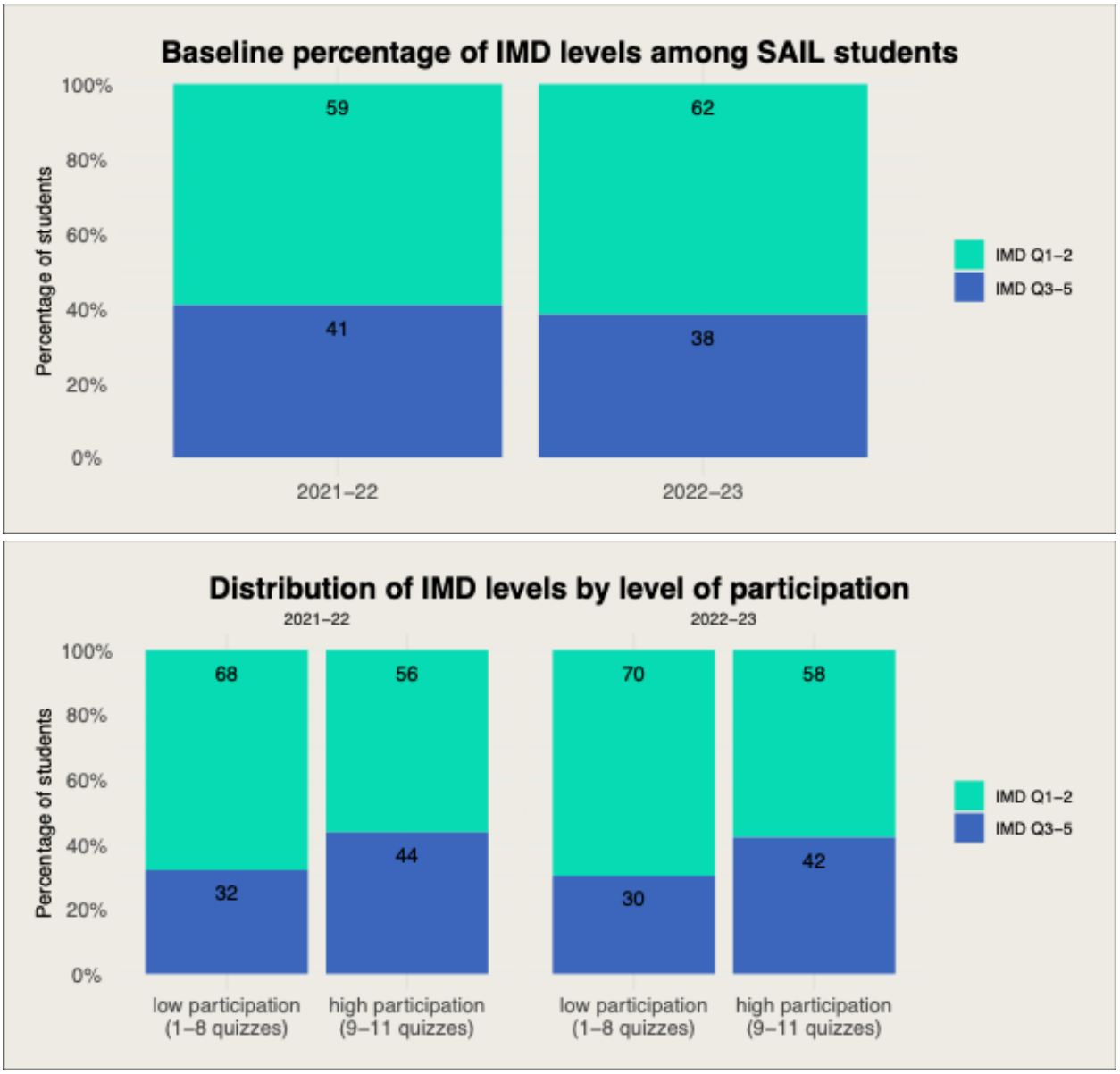

Figure 8. Percentage of students from areas of low (IMD quintiles 1-2) and high (IMD quintiles 3-5) levels of deprivation (a proxy for socioeconomic status) among all SAIL participants (top panel) and split by the level of participation in SAIL (bottom panel).

The data indicates that:

1) There is no difference in rates of participation by ethnicity of students in both academic years.
2) There is a clear difference between low- and high-participating students in terms of entry qualifications. In both 2021-22 and 2022-23, there are more A-level students in the high-participation category, and more BTEC students in the low participation category. In both years, the proportion of low participating students with BTEC background is higher than among the overall population of SAIL

students. Conversely, the proportion of low participating students with A level background is lower than among the overall population of SAIL students, although this difference is not as strong in 2022-23.

3) There is also a change in IMD levels by degree of participation compared to baseline proportions, with students from lower IMD levels overrepresented in the low participation category.

The findings #2 and #3 above are potentially concerning. As detailed in the SAIL theory of change, one of the motivations for introducing SAIL and one of its intended effects is development of effective study habits among the students, an element of which is a regular revision of topics and materials covered in class. This, it was hoped, would particularly benefit students who arrive at the university without being accustomed to the type and level of work required for successful academic performance.

At Huddersfield, we have been aware for some time that students arriving with BTEC-type entry qualifications on average perform worse than A-level students, in terms of module marks and degree outcomes. Initiatives such as SAIL could potentially bridge this gap. The plots above suggest, however, that even a process with a low barrier to entry and a significant incentive to completion does not prevent differential participation rate depending on entry qualifications and socioeconomic status of the students. This observation calls for an in-depth investigation of the causes of this participation differential and redesign of the approach to encourage participation among all groups of students.

Having said that, it needs to be emphasised that the very small window this quantitative data provides is insufficient to provide an adequate understanding of the processes and motivations in play even for such a simple intervention. We nevertheless hope that this example evaluation provides multiple points of entry for further inquiry that must be less limited and must involve qualitative information from students and staff themselves.

5. Discussion

How was the SAIL initiative implemented?

The SAIL programme, implemented entirely through the functionality of the virtual learning environment (VLE) platform, has worked well, although not consistently throughout all quizzes. About half of the quizzes were set up to be available for one week only and to contain questions drawn randomly from a pool (such that each student would receive a different set of questions) – as intended. The rest was either available for longer than a week (usually due to a missing availability limit on a quiz) and presented the same set of questions to each student. We do not believe these differences would affect the SAIL goals (and we have demonstrated in this IPE that it did not affect participation in SAIL), but, importantly, these technical issues can be resolved through the VLE in the future editions of the programme.

How did the students engage with SAIL initiative?

The engagement with SAIL by the students was high. Most students completed most of the quizzes, week to week, with high median marks (80%) for each quiz. This demonstrated that SAIL provides a good incentive for students to participate and that most students are able to provide correct answers to the quizzes. Nevertheless, a more in-depth look at patterns of engagement for different groups of students found that students with BTEC-like entry qualifications and students from lower socioeconomic background tend to be overrepresented among the students who completed between 1-8 quizzes compared to overall SAIL participants and that these students also achieve lower SAIL scores compared to students from the high participation group. Still, the fact that “low engaging group” contains students who completed up to 8 quizzes indicates a high participation rate.

These findings suggest that the SAIL approach on its own does not eliminate the differential attainment gap we observe at Huddersfield for students with different entry qualifications and with different socio-economic backgrounds. The detailed explanation of the mechanisms of student (and staff) engagement in play in the SAIL programme requires, however, further research – both quantitative and qualitative, particularly on following the pattern of engagement of individual students over time, and querying them on their circumstances and motivations.

5.1. Limitations of the research.

5.1.1. Evaluation of student engagement

One of the main reasons to introduce SAIL was to improve student engagement. In the IE, we found that attendance, one of the most commonly used measures of student engagement, does not differ between SAIL and non-SAIL populations. SAIL also does not differentiate module marks in the two groups. The straightforward interpretation of this finding is that SAIL does not affect these measures of student engagement. However, it also prompts a question whether a single, quantitative measure of enforced participation – attendance (see below) – may be a poor measure of a meaningful academic engagement. Rather, it reflects what we may call a *performative student engagement*: students with good attendance and assignment submission records, yet for whom we are unable to detect an improvement in academic outcomes, compared to students without these characteristics.

An additional complication in interpreting the IE finding is that attendance at the University of Huddersfield is mandated and monitored, with student support being activated for students who attend their timetabled classes below a specified threshold. Therefore, the observed attendance patterns may be driven mainly by an avoidance of formal low-attendance meeting rather than a desire and an active effort to learn. Even though there is a positive relationship at Huddersfield between attendance levels and module marks, this enforced attendance may obscure the differentiation potential of this measure of engagement between SAIL and non-SAIL students.

5.1.1.1. What is engagement?

The concept of a meaningful student engagement is, of course, difficult to define. If we assume, for the sake of this discussion, that it involves “actions that students do to become competent in an academic discipline”, they likely involve non-quantitative factors such as individual approach to study and revision, perceived difficulty (or lack of thereof) of the assignments, motivation to come to and stay at the university, motivation of achieve a specific outcome at or upon graduation from the university, personal circumstances and environmental factors (work, housing, financial issues, family matters, etc). There is currently no quantitative measure that can be deployed at scale to survey students that captures the complexity of what differentiates “good” students from the rest. Even the concept of what/who is a good student is complex: there are multiple facets of being a student and different ideas of what constitutes “good”, from the perspective of students, staff and their wider environment. As an example of the limitation of the data presented here, we have records of students taking the quizzes, but we do not have records of the students checking and acting upon their feedback from each quiz (including the correct answers to questions). Arguably, engaging with feedback is an important element of a meaningful academic practice that we would like to engender in students, yet it remains unexplored in this approach.

In our view, the experience of IE clearly demonstrates the limitation of reducing engagement to a single value for each student, and articulates a need to build a much richer understanding of meaningful student engagement, one that will not exclusively rely on quantitative, de-personified data but that will include a host of qualitative data such as in-depth structure interviews with students.

5.1.2. Other limitations

Mirroring the latter concern, an important limitation of the SAIL evaluation process is a lack of quantitative *or* qualitative information on the staff's engagement with the intervention. We have no data that could help us investigate staff's attitudes to SAIL, their perception of the workload involved in its preparation and how they interacted with the entire process. For example, we don't know how often and how many members of staff regularly reviewer the quizzes' results, how often they and how many of them acted on it by organising extra tutorial sessions and whether they noticed any change in students' attitudes and approaches to the learning process over the course of the term that could be attributed to SAIL. Collecting these pieces of information from "the other side" of an institutional intervention would contribute a valuable insight into factors affecting the process' effectiveness.

Last but not least, both the IE and IPE require considerable time and human resources from the data perspective. For example, demographic, attendance and VLE data come from separate and independent systems under care of separate teams at the university; these systems have different levels of data curation and use different formatting of student identifiers. To our knowledge, similar datasets at other institutions are rarely centralised. Collection, standardisation and curation of such data requires experience and expertise that departmental-level teams, who would be best placed to understand the data, may not always possess. If we argue – as we do – that an even more sophisticated data view of the implementation is necessary for any intervention (including qualitative data), the in-time data collation and analysis becomes a critical competency that must be developed at all levels of institutional hierarchy.

5.1.3. Internal validity

The SAIL approach is straightforwardly transferrable across the institution. The low-stakes summative assessment strategy can be applied to different courses, disciplines and institutions. The key potential concern is workload, in preparation of the database of questions and in marking the answers. Adjacent to this issue, we note that some disciplines, particularly those that rely on assessing writing skills and those where

assignments require production of physical artifacts (e.g. in subjects such as humanities, arts and architecture), generation of easily and automatically marked questions may require a bigger effort. STEM disciplines, or any discipline where answers can be calculated or are otherwise easily defined (e.g. names and details of processes or structures in life sciences) can be implemented relatively easily. The SAIL quizzes implemented in CEET provided a good variety of questions throughout the term and across the modules.

This IPE also demonstrated that students, likely motivated by a relatively high proportion of overall marks available, overwhelmingly participate in the initiative and maintain their involvement over the course or the entire term. We did not observe differences pointing to overrepresentation of some groups of students among low or high scorers in SAIL or among students who submit a small or large number of SAIL quizzes.

5.1.4. External validity

Overall, the evaluation presented in the IE and this IPE is in principle a valid approach to assessing the impact of SAIL, but given a very limited scope, where outcomes are reduced to attendance and only quantitative data is used in the analysis, it has low discriminatory power. This highlights the need for and importance of surfacing the complexity of hypothesised impact of an intervention as well as sophisticated means of collecting quantitative and qualitative information about student and staff behaviour, academic results and their metacognitive processes, to fully understand the mechanisms driving students' and staff's participation and engagement with any intervention.