Research protocol
Leicester Race Equality Gaps Analysis

Author: Yihan Xu, Giulia Tagliferri

QA: James Lawrence at BIT

BIT TP number: 2021068

VERSION	DATE	REASON FOR REVISION/NOTES
Any changes to the design to be agreed between the implementation partner(s), evaluator and TASO. Note any agreed changes in the table below.		
1.3	11.2.22	Amended the order and in-text citation of tables to make the content consistent
1.2 [*original*]	2.2.22	
Pre-registration		This design has been pre-registered on the Open Science Framework.

The QA rating system is based on the Evaluation Security tool presented in the TASO Monitoring and Evaluation Framework.[1]

QA	Comments	Rating (out of 5)
Design	Matched Difference in Differences	4
Sample size	Expected to be 2772 observations	4
Outcome measure	Administrative data on attainment	5
Attrition	No attrition due to use of administrative data	5
Validity	Adequate counterfactuals identified	4

Overall		

1. Summary

Background
The Centre for Transforming Access and Student Outcomes in Higher Education (henceforth TASO) has funded the University of Leicester (henceforth Leicester) and commissioned the Behavioural Insights Team (henceforth BIT) to evaluate the impact of their "Decolonising the Curriculum Toolkit" (a resource for staff that provides clear and concise guidelines on how to make their curriculum more racially inclusive) on reducing attainment gaps between Black, Asian and Minority Ethnic (BAME) students and White students.

Aims
To evaluate whether and to what extent Leicester's 'Decolonising the Curriculum Toolkit' reduced the attainment gaps between BAME and white students.

Intervention
The "Decolonising the Curriculum Toolkit" is a two-page resource for staff that provides clear and concise guidelines on how to make module content, assessment and practice more racially inclusive and relatable for all students. The toolkit was piloted across all modules in the Sociology BA course in the 2020/21 academic year.

Design
The study is a matched difference-in-differences with repeated cross-sections. We will compare attainment trends among the modules that implemented the "Decolonising the Curriculum Toolkit" (treatment modules) with similar comparator modules that didn't implement the initiative.

Outcome measures
The primary outcome measure is the module-level average attainment, and it is defined as the percentile rank of the final module mark.

Analyses
The primary analysis consists of a difference-in-differences regression, comparing module marks before and after the academic year 2019-20 (the year that curriculum reform took place) between reformed vs. matched unreformed modules. It will focus on BAME students only. The secondary analysis will repeat the primary analysis for White students. Additional descriptive line charts will be made to illustrate the change in attainment gaps of reformed vs. comparator modules before and after the "Decolonising the Curriculum Toolkit" was implemented.

2. Background

This research is part of a TASO-funded project to evaluate the impact of universities' efforts to reform curricula as a means of reducing racial equality gaps in student outcomes.

2.1 Funding sources

This research is funded by TASO. TASO has funded a research assistant in each of the two partner universities (Kent and Leicester) to evaluate the impact of diversifying curricula and has commissioned BIT to deliver the quantitative evaluation of both universities

2.2 Team, role, and responsibility

Table 1 presents an overview of the project team. BIT's researchers will lead the design and analysis of the quasi-experimental analysis, leading in the development of the trial protocol and the analysis report. They will use GDPR-compliant administrative data provided by Leicester to investigate the effect of the curriculum reform on the attainment trends of BAME and white students.

TASO's colleagues will review BIT's proposed analytical approaches and provide comments as needed. TASO will also help facilitate the collaboration between BIT and Leicester in data and knowledge transfer related to this project (e.g. signing data sharing/processing agreement).

Leicester colleagues will share background information with BIT as well as sharing GDPR-compliant individual-level module data. In addition, they will also help BIT address project or data related questions as needed.

Table 1. Core project team, roles and responsibilities

Organisation	Name	Role and responsibilities
BIT	Dr Giulia Tagliaferri	Research lead
BIT	Dr Yihan Xu	Research analyst
BIT	Dr Alex Sutherland	Evaluation quality assurance
BIT	James Lawrence	Evaluation Supervisor and quality assurance
TASO	Sarah Chappell	Project lead
TASO	Dr Helen Lawson	Research lead
Leicester	Dr Paul Campbell	Partner lead
Leicester	Dr Hannah Grosvenor	Partner co-investigator
Leicester	John Hurst	Partner data curator
Leicester	Clare Amess	Partner data curator

3. Aims

3.1 Research questions

The primary research question:
How did Leicester's 'Decolonising the Curriculum Toolkit' affect the attainment of BAME students?

The secondary research question:
How did Leicester's 'Decolonising the Curriculum Toolkit' affect the attainment of White students?

The exploratory research question:
How did Leicester's 'Decolonising the Curriculum Toolkit' affect the attainment gap between White and BAME students?

3.2 Research hypotheses
We hypothesise that undergraduate Sociology core modules that diversified their curricula will have smaller White/BAME attainment gaps post-intervention than comparator modules that did not diversify their curricula.

3.3 Rationale for choosing comparators
Comparator modules were chosen to establish plausible counterfactuals, for participation in the 'Decolonising the Curriculum Toolkit' initiative was voluntary for module instructors, therefore module reformation could not be (nor could be considered) randomly assigned. See Section 5.4 for details on matching methodology.

4. Intervention

4.1. Overview of the 'Decolonising the Curriculum Toolkit'

In response to a below 80% student satisfaction score in the NSS in 2018/19 for the Sociology Degree at the University of Leicester, Dr Paul Campbell designed and developed the 'Decolonising the Curriculum Toolkit' – a two-page resource for staff that provides clear and concise guidelines on how to make module content, assessment and practice more racially inclusive and relatable for all students. The toolkit has the following objectives:

- Provide a set of practical guidelines to help make teaching practice more inclusive and more responsive to the student body.
- Help academics to reflect on their programmes of study, modules and cultural practices, with the aim of making the sociology curriculum more engaging and better connected to all the students it serves, educates, and seeks to inspire.
- Improve student satisfaction and attainment

The toolkit was designed to improve the racial literacy of staff by providing a short-hand and accessible resource which staff can work through in their own time and with little formal training. It purposely does not provide an exhaustive and prescriptive set of instructions, but by providing a host of conversational questions it prompts more meaningful reflection and strategies on how to improve their practice and racial literacy in ways which they can incorporate best into their practice. This is much more meaningful and less didactic approach for improving our teaching staffs' knowledge of race and the way it works; its place within their disciplines; its place within their students' lives and in the lives of the people who students will work alongside; and for how to begin the process of disrupting this.

The toolkit provides teaching-staff with the tools for critical reflection with regards to race. That is to help them be able to better recognise, dismantle and guard against the ways in which course content, assessment and practice can marginalise and benefit students from certain backgrounds and contribute to barriers, lower satisfaction and the award gap for some.

4.2. Implementation of the 'Decolonising the Curriculum Toolkit'

The toolkit was piloted across three Sociology modules during the 2019/20 academic year. Data (in the form of staff testimonies) demonstrated that in terms of impact for fostering inclusive practice, the toolkit had been extremely effective in aiding teaching-staff to reflect on the racial inequities that might exist within their pedagogical practice or content, and improve individual's confidence to meaningfully reflect on, and take ownership of, the decolonizing process.

Against these early indicators for success, the toolkit was piloted across all modules in the Sociology BA course in the 2020/21 academic year. However, Leicester does not operate a mandatory approach to inclusion interventions which has meant that the full usefulness of the toolkit is in part determined by the willingness and time for staff to engage with the intervention.

At the end of the academic year, Dr Paul Campbell asked module convenors to give a rank score out of 10 for the level of engagement with the 'Decolonising the Curriculum Toolkit' when devising, planning and or delivering content for their module during the 2020/21 academic year. Values given were from '0', which did not engage with the toolkit at all, to '10', engaged with the toolkit in its entirety.

5. Design

BIT will use a matched difference-in-differences approach to evaluate the impact of the curricula reform initiative, where comparator modules will be matched to reformed modules on pre-intervention module characteristics. BIT will then compare the pre-intervention and post-intervention trend of students' attainment among the reformed modules with comparator modules that didn't reform their curricula.

5.1 Module inclusion & exclusion criteria

The treated modules will be selected from the Sociology programme, whereas the comparator modules will be chosen from a pool of unreformed modules from three other programmes (Chemistry, Criminology, and Geography) that had characteristics most similar to that of the Sociology programme, as well as unreformed modules from the Sociology programme.

To maximise the comparability of modules, we will only include modules for further analysis if they met the following criteria:
- The module is not a graduate level-7 module
- The module credit is between 10~45 credits as modules with more than 45 credits typically involve a dissertation, and modules with fewer than 10 credits did not have sufficient exposure to and stake of the interventions
- Have 10 or more students enrolled in 2021
- Have at least 3 year of pre-intervention attainment data

A total of 77 modules met the above criteria. Among these, 14 were reformed in 2020/21, whilst 63 are potential comparator modules that were not reformed at any point of time (see **Table 2.1** for details).

Table 2.1 Number and characteristics of eligible modules by programme

Programme	Reformed status	Number of eligible modules	Compulsory modules (n, %)	Advanced modules (n, %)	Average number of students enrolled in 2021 (mean, SD)
Sociology	Yes	14	6 (42.9%)	5 (35.7%)	40.9 (20.7)
Chemistry	No	13	5 (38.5%)	12 (92.3%)	57.8 (50.3)
Criminology	No	14	7 (50.0%)	2 (14.3%)	86.4 (42.7)
Geography	No.	36	14 (38.9%)	16 (44.4%)	32.5 (18.0)

5.2 Module reformed status

For Sociology modules to be considered as reformed, their intervention intensity score (as judged by the module convenor's engagement with the toolkit) should be deemed as 4 (inclusive) or higher (out of a scale of 10). According to an assessment made by Dr Paul Campbell from Leicester University:

- Among the 14 Sociology modules, 4 modules ("SY1021","SY2078", "SY2093", "SY3095") were rated as having an intervention intensity score of lower than

4. Those four modules were no longer counted as reformed, and together with the other 63 unreformed modules, formed a pool of comparator modules;

- Two modules ("SY3092" and "SY3093") that had an unknown intervention intensity score were also excluded, leaving 8 modules as reformed;

In sum, a total of 75 eligible modules remained for further analysis. The overall characteristics of those modules are presented in **Table 2.2**.

Overall speaking, though the general characteristics of the reformed modules and the pool of comparator modules are somewhat comparable, they were not sufficiently similar as shown in **Table 2.2**, therefore matching is needed to identify a more robust counterfactual (comparator group).

Table 2.2 Number of eligible modules by reformed status

	Reformed modules	Pool of comparator modules
Number of modules	8	67
Compulsory modules (n, %)	4 (50%)	28 (41.8%)
Advanced modules (n, %)	2 (25%)	31 (46.3%)
Courses (n, %)	Sociology: (8, 100%)	Sociology: 4 (6.0%) Chemistry: 13 (19.4%) Criminology: 14 (20.9%) Geography: 36 (53.7%)
Average number of students enrolled between 2018 to 2020 (mean, SD)	43.4 (19.9)	50.5 (28.5)
Average proportion of BAME students between 2018 to 2020 (mean, SD)	57.0 (5.6)	31.8 (16.3)
Average mark in percentile rank between 2018 to 2020 (mean, SD)	40.2 (4.96)	44.3 (8.01)

5.3 Module-matching procedure and results

The comparator modules will be selected from the pool of eligible comparator modules. They will be matched based on how similar they were to the reformed modules pre-intervention in the following characteristics:

- Whether module is compulsory or elective
- Whether module level is entry level (level 2 or below) or advanced level (level 3 and 4)
- Average number of enrolled students from 2018 to 2020
- Average percentage of BAME students from 2018 to 2020

- Average attainment (percentile rank of the final module mark) among BAME students from 2018 to 2020

The matching was done using the R package MatchIt[2]. Each reformed module was matched based on the above-mentioned matching criteria.

The modules were assigned a propensity score, indicating the fitted likelihood that the module was reformed given its characteristics. Matching was done on a 1:1 basis, without replacement, as this is a conservative matching method which is also intuitive to interpret. The matching was done separately for each reformed module. **Table 3** presents the propensity scores of the reformed modules pairing with eight comparator modules that had the closest propensity scores.

Table 3 Propensity scores of reformed vs. comparator modules

| Matched pair	Reformed module	Propensity score	Comparator module[3]	Propensity score
Pair 1	SY1002	0.907	CR2026	0.741
Pair 2	SY2089	0.833	SY2078	0.560
Pair 3	SY2090	0.649	SY2093	0.413
Pair 4	SY2091	0.541	SY3095	0.348
Pair 5	SY2092	0.316	CR3020	0.281
Pair 6	SY2093	0.271	CH0061	0.187
Pair 7	SY2094	0.115	CR2023	0.117
Pair 8	SY2095	0.080	CR1001	0.080

5.4 Visual inspection of the parallel trend assumption

We calculated BAME students' module-level weighted average attainment of the reformed and comparator modules up to 3 years prior to intervention. We then plotted the parallel trends in **Figure 2**. It appears that the trends were parallel from 2019 to 2020, though the trend seemed less parallel in 2018. In the next section, we will specify how we test the parallel trend assumption formally.

Figure 2. Trends in weighted average module mark before intervention

---

[2] Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199–236. doi: 10.1093/pan/mpl013

[3] CR is shorted from Criminology, SY for Sociology, CH for chemistry, GY for Geology
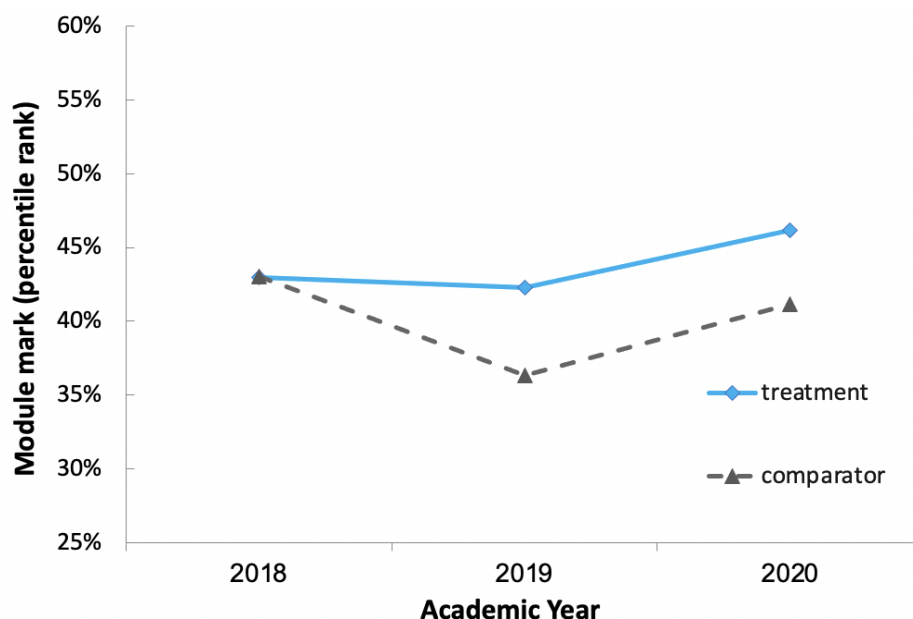
5.5 Formal testing of the parallel trend assumption

We used a similar regression specification as the main regression (see **Section 11**) to test whether the pre-intervention trends of module mark (percentile rank) between treatment and comparator modules were parallel.

The regression outputs (using data up to 2019-20) showed that although the trend was statistically significantly less parallel in 2018, the trends in 2019 and 2020 were not statistically significantly different. As a result, we think the reformed modules and the matched modules had an adequately parallel trend before the intervention.

6. Outcome measures

This study only has one outcome measure, and it's listed in the table below.

Table 4. Outcome measures

Outcome measure	Data to be collected	Point of collection
Primary outcome: Final module mark in percentile ranking	Raw final module grades for all students of the modules listed in Table 2.2 from academic year 2017-18 to 2020-21. Data will be anonymised before sharing.	The data is routinely collected by Leicester and will be provided (sent in two batches, in Aug and Nov 2021) by Leicester once the BIT-TASO data processing agreement and the Leicester-TASO data sharing agreement are signed.

We will use percentile rank of module mark instead of the raw mark as the outcome measure for the following reasons:

- Percentile rank is less susceptible to trend, e.g., grade inflation
- Percentile rank is also less susceptible to course instructors' grading style (some instructors' 70 might be equivalent to others' 60) as the highest value (whether it's 70 or 90) will always be standardised to 100 and the lowest value will be standardised to zero, making between-module difference more objective and comparable
- Percentile is more intuitive to interpret, e.g., if BAME students' average percentile rank is 50%, it indicates zero White-BAME gap.
- Lower risk of de-identification of module instructors (See Section 12.2 for details)

On the other hand, using raw marks as the outcome measure does have some benefits as the OfS uses this metric to calculate awarding (% of students achieving first/second class honour) gaps. We acknowledge that our primary approach differs from the OfS approach, however, we think overall the benefits outweigh the risks. Furthermore, in order for the output to be better comparable to other reports in this area, we will visualise the degree awarding gap using both percentile rank and percentage of students receiving either an upper second class or a first class in the modules (see Section 11).

6.2 Interpretation of the outcome measure
Although the theoretical range of both the raw module mark and the percentile rank of module mark is from 0 to 100, in practice, the range of the latter is likely to be much wider than the former, because few instructors seldom give marks higher than 80 or lower than 40. Descriptive analysis (using data up to 2019-20) shows that the mean raw mark was 60.98 for White students and 56.24 for BAME students, whereas the mean percentile rank was 52.84 for White students and 42.73 for BAME students. Thus the gap might seem wider if we use percentile rank, thus should be interpreted with caution.

As stated above, we will visualise the attainment gap for White and BAME students using both percentile rank and the percentage of students who achieved an upper second class and first class so that this is comparable.

7. Sample selection

7.1 Study settings

The curriculum decolonisation initiative was piloted among cohorts enrolled in Leicester's Sociology BA course. The Sociology BA is a full-time campus-based course, which has approximately 219 students of which 47.5% (104) self-describe as BAME.

7.2 Inclusion and exclusion criteria

The sample comprises BAME and White students' final module marks (in percentile rank) of matched modules from four programmes (Sociology, Chemistry, Criminology, and Geography) in the following academic years: 2017-18, 2018-19, 2019-20, and 2020-21.

7.2.1 For modules

A total of 8 pairs (see Table 3) of successfully-paired modules were included for final analysis met the following criteria as specified in **Section 5.1**:
- The module is not a graduate level-7 module
- The module credit is between 10~45 credits, as modules with more than 45 credits typically involve a dissertation, and modules with fewer than 10 credits did not have sufficient exposure to and stake of the interventions
- Have 10 or more students enrolled in 2021
- Have at least 3 year of pre-reformed attainment data

7.2.1 For students

To minimise potential selection bias, within the included modules, we excluded module mark records of students whose:
- Ethnicity is unknown
- Fee payment status is other than the UK (this is consistent with the approach of the OfS. Furthermore, UK fee payers will likely have gone through the UK education system, so this exclusion reduces spurious factors).

7.3 Sample size

After applying the inclusion and exclusion criteria, we are left with the following sample sizes (See **Table 5.1** for the total sample size and **Table 5.2** for the subsample of BAME students).

In total, we had 2772 observations of module mark records from 2017-18 to 2020-21 and on average, 53.2% of them belonged to BAME students. Among the total sample, about 23.7% (656 out of 2772) of the records took place post-intervention.

Table 5.1 Total sample size (including both BAME and white students)

Academic year	Reformed Modules		Comparator Modules	Overall	
	un-reformed	reformed	un-reformed	un-reformed	reformed
2017-18	438	-	385	823	-
2018-19	381	-	422	803	-

	Reformed Modules		Comparator Modules	Overall	
2019-20	220	-	270	490	-
2020-21	-	321	335	335	321
Total	1039	321	1412	2451	321

Table 5.2 Total sample size for BAME students

Academic year	Reformed Modules		Comparator Modules	Overall	
	un-reformed	reformed	un-reformed	un-reformed	reformed
2017-18	243	-	204	447	-
2018-19	217	-	223	440	-
2019-20	109	-	150	259	-
2020-21	-	173	156	156	173
Total	569	173	733	1302	173

8. Data collection

Data was collected in two batches. The data from academic year 2017-18 to 2019-20 had already been collected and was part of the University of Leicester's institutional dataset before the kick-off meeting. The data from the academic year 2020-21 was collected by Leicester over the summer of 2021 and then provided to BIT in November 2021.

For the purpose of this study, BIT's researchers received pseudonymised data only. Main data items collected are listed in **Table 6**. The data were transferred via encrypted method and also stored encrypted on BIT servers and access will be restricted to researchers directly involved in this project. All researchers' laptops have anti-virus applications installed and encrypted hard-drives to protect data stored locally.

The code for cleaning and analysis were all quality assured at BIT. Further details of data management procedures are specified in the data sharing agreement.

Table 6. Data collection

Data item	Timeframe	Collector	Data collection status
Module reformed status, Module attainment, Student covariates, Module characteristics	Academic year 2017-18 to 2019-20	Leicester	Shared with BIT in Aug 2021
Module reformed status and module reformed intensity, Module attainment, Student covariates, Module characteristics,	Academic year 2020-21	Leicester	Shared with BIT in Nov 2021

9. Procedure

The high-level project plan is presented in **Table 7**.

Table 7. Project plan for deliverables

Timeframe	Action	Owner
June 2021	Kick off meetings to clarify roles & responsibilities	BIT, TASO, Leicester
Sep ~ Nov 2021	Pre-Analysis (part 1) - Agee on DSAs (June 2021) - Define scoping of research design and options (Nov	BIT, TASO, Leicester
Nov 2021	Break point and presentation of scoping ahead of full protocol development	BIT
Jan 2022	Pre-Analysis (part 2) - Draft trial protocol - Agree on draft trial protocol	BIT, TASO
End of Jan - Feb 2022	D-in-D Analysis	BIT
Feb 2022	Draft analysis report and orally present the analysis results	BIT

10. Power calculations

We used the following information to estimate MDES:

- Significance level: 0.05
- Power: 0.8
- Expected sample size (see **Table 5.1** and **Table 5.2**)
- Standard error of the coefficient for treatment effects using the data set up to 2019-20: 3.54
- Standard deviation of module mark (in percentile ranking): 28.43

- Constant used to estimate MDES: qnorm(0.975) + qnorm(0.8) = 2.80

As a result, we calculated the MDES (in percentile ranking) as 2.80*3.54 = 9.91. In other words, **we're powered to detect a difference of 9.91 percentile rank of final module marks before vs. after intervention among BAME students**. If expressed in the unit of Cohen's D, the MDES is 9.91/28.43 = 0.35, i.e. we're powered to detect a medium difference. It's worth noting that this is likely to be an under-estimation of the MDES as we haven't included the 2020-21 data when estimating the standard error.

11. Analytical strategy

The primary analysis will focus on BAME students only, including data from the academic year 2017-18 to 2020-21. The analysis will be a difference-in-difference regression with three years of pre-intervention data points and one year of post-intervention data points. The OLS regression model is specified as follows:

$$Y_{imt} = \beta_0 + \delta \; PostInvervention_t \times EverTreated_m + \beta_1 Time_t +$$
$$\beta_2 EverTreated_{mt} + \beta_3 Gender_i + \beta_4 MatchedPair_m + \beta_5 ModuleLevel_m +$$
$$\epsilon_{imt}$$

Where:

- Y_i denotes the final module mark (in percentile rank) of individual i of module m in academic year t
- β_0 is the constant
- δ is the causal effect of interest, representing the difference in attainment trend for reformed modules in the post-treatment period(s). $PostInvervention_{mt} = 1$ if by academic year t, the intervention had taken place for the reformed module m and its matched module; $PostInvervention_{mt} = 0$ if the intervention had not. $EverTreated_m = 1$ if module m was ever reformed; $EverTreated_m = 0$ if module m was never reformed.
- $Time_t$ is a set of dummies that take value from 2017-18 to 2020-21.
- $Gender_i$ denotes the gender of participant i gender (0 = female; 1 = male).
- $MatchedPair_m$ is a set of dummies that denotes one of eight pairs of modules successfully matched by propensity scores based on module characteristics.
- $ModuleLevel_m$ is a set of dummies that denotes whether the module is elementary or advanced.
- ϵ_{imt} is an individual-level error term.

A descriptive exploratory analysis will deduct the attainment gap between White and BAME students from the primary and secondary analysis, and the race gap results will be visualised using line charts.

12. Ethical considerations

12.1 Data protection
This study has high standards of data protection. TASO and BIT have cooperated closely with Leicester's data protection officer to ensure that this study complies with all the data protection protocols. It only uses pseudo-anonymised individual-level and module-level data. No identifiable information will be elicited from students and module instructors. Because data is fully anonymised and shared using encrypted methods, there is little risk to the students and module instructors involved in this study.

12.2 Risk of de-anonymisation
It is possible that colleagues within the participating school might be able to infer the treated modules by reading the number of enrolled students or the attainment gaps. This may cause some self-consciousness for module instructors if the modules they taught deteriorated over the years compared to other modules. To mitigate this risk, we have standardised module marks by converting the raw mark into percentile rank so that it's less straightforward to make inferences about. We have also aggregated the results at "treatment" vs. "comparator" level, further lowering the risk of individual modules being identified.

12.3 Consent
Consent is addressed through the university's standard data protection agreement with students, rather than separate consent for this study. The data will be retained for 12 months after the delivery of the final report. The agreed date for the deletion of all evaluation and research data shall be March 31st, 2023.

13. Risks

Table 8. Potential risks and mitigation strategies

Risk	Mitigation strategy	Risk owner
Lower student engagement with reading materials in the academic year 2019-20 and 2020-21 due to the remote teaching mode during the pandemic	BIT will include fixed effect for academic year to control for potential effects	BIT

Data delay due to DSA/DPA signing	BIT and Leicester team will hold regular team meetings to engage relevant stakeholders	Leicester
Fail to identify matched modules	BIT will use available data to explore multiple matching strategies and proceed with the most promising one	BIT