Analysis report

University of Leicester Impact Evaluation Analysis

November 2022

Author: Dr Yihan Xu (BIT), Dr Giulia Tagliaferri (BIT)

QA: Dr Patrick Taylor (BIT)

The full protocol for this study can be found on the TASO website

The study was pre-registered on the Open Science Framework

1. Summary

Background

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

Aims

To evaluate how Leicester's 'Decolonising the Curriculum Toolkit' affected the attainment of BAME and White students as well as the racial awarding gap.

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 teaching practice more racially inclusive and relatable for all students. The toolkit was piloted across the Sociology BA course in the 2020/21 academic year.

Design

This is a matched difference-in-differences study with repeated cross-sections. The analysis compares students' attainment trends in the modules that implemented the "Decolonising the Curriculum Toolkit" (treatment modules) with that of similarly comparable modules that did not implement the initiative.

Outcome measures

The primary outcome measure is a student's module-level 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 focuses on BAME students only. The secondary analysis repeats the primary analysis for White students. Additional descriptive line charts have been made to illustrate how the awarding gaps of reformed vs. comparator modules changed since the "Decolonising the Curriculum Toolkit" was implemented.

Results

Overall, this impact evaluation suggests that the “Decolonising the Curriculum Toolkit" might have had a negative impact on both BAME and White students’ attainment among Sociology students at the University of Leicester. The estimated treatment effect was significantly negative among BAME students, **-6.63** percentiles, **95% CI [-13.23, -0.03]**, p = 0.05. It was directionally negative (though not significant at the 5% level) among White students, **-3.07 percentiles**, **95% CI [-9.79, 3.64]**, p = 0.37.

Findings from the exploratory analysis suggest that intervention did not affect the racial awarding gap.

Conclusions

In light of the above discussion, we do not recommend rolling out this toolkit (in its current form) before conducting a closer examination of how the toolkit was delivered by teachers and received by students. We believe the implementation and process evaluation (IPE) led by Leicester may shed light on what might have caused this and help contextualise these effects. If findings from the IPE suggest that there is evidence of promise in how teaching staff and BAME students might benefit from the reforms, we would recommend refining the toolkit based on the IPE and then conducting further impact evaluation of the intervention, with a larger sample and over a longer time period.

2. Introduction

2.1. 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 associate at the University of Leicester to support on the evaluation and has commissioned BIT to deliver the quantitative (impact) evaluation.

2.2 Team, role, and responsibility

Table 1 presents an overview of the project team. TASO instructed BIT to propose the details of a Differences-in-Differences design to answer the research question at hand, using administrative data provided by Leicester. Leicester colleagues shared background information with BIT and helped BIT address project or data related questions as needed. In addition, they shared GDPR-compliant individual-level module data with BIT and led in the drafting of the background and intervention sections of the trial protocol and the analysis report.

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 (trial protocol stage)
BIT	Dr Patrick Taylor	Quality assurance (analysis report stage)
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

Leicester	Dr Ashjan Ajour	Research associate

2.2. Aims

2.2.1 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 awarding gap between White and BAME students?

2.2.2 Research hypotheses

We hypothesised that undergraduate Sociology core modules that engaged with the 'Decolonising the curriculum toolkit' in the 2020/2021 academic year would have a smaller White/BAME awarding gap post-intervention than comparator modules that did not engage with the toolkit.

2.2.3 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 3.1.3** for details on matching methodology.

2.3. Intervention

2.3.1 Overview of the 'Decolonising the Curriculum Toolkit'

University of Leicester's 'Decolonising the Curriculum Toolkit' (DCT; see Appendix 3) is a two-page resource for staff that provides guidelines on how to make module content, assessment and practice more racially inclusive and relatable for all students. The toolkit was designed to improve the racial literacy of staff by providing a short and accessible resource which staff can work through in their own time and with little formal training. It deliberately does not provide an exhaustive and prescriptive set of

instructions, but through a host of conversational questions, prompts more meaningful reflection and strategies on how to improve racial literacy and best incorporate it into practice. The toolkit also provides teaching staff with the tools for critical reflection with regard to race to help them better recognise, dismantle and guard against how course content, assessment and practice can marginalise or benefit students from certain backgrounds and contribute to barriers, lower satisfaction and the awarding gap.

The intervention was piloted across all modules in the Sociology BA course in the 2020–21 academic year. The resource was made available to all staff via the university intranet; however, it was not mandated and there were no formal requirements for engagement or accountability placed on staff to operationalise the toolkit. The assumption was that the guidance provided would ensure consistent levels of adaptation to content across all taught modules. This position was based on the DCT initial pilot in 2020, where the staff surveyed reported that they found the toolkit easy to follow.

Consequently, it was anticipated that consistent levels of change would be seen in content across all modules within the undergraduate degree in which the intervention was tested. Typically, it was envisaged that this would manifest in the following ways:

- Levels of diversity and pluralising of narratives/viewpoints in reading lists (minimum of 20% of weekly core readings),
- An audit and inclusion of racially inclusive imagery across all module content,
- A significant increase in the explicit opportunities offered to students to relate taught content and assessments to their own lived context or biographies.

2.3.2 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) suggested 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 promise, the toolkit was piloted across all modules in the Sociology BA course in the 2020/21 academic year. However, Leicester did not mandate the inclusion interventions which means that the staff can determine whether and to which extent they can engage with the toolkit.

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 to 10 (0 = did not engage with the toolkit at all;10 = engaged with the toolkit in its entirety).

3. Methods

3.1. Design

BIT used a matched difference-in-differences approach to evaluate the impact of the curricula reform initiative, where comparator modules were matched to reformed modules on pre-intervention module characteristics. BIT then compared the pre-intervention and post-intervention trend of students' attainment among the reformed modules with comparator modules that did not reform their curricula.

3.1.1 Module inclusion & exclusion criteria

The treated modules were selected from the Sociology programme, whereas the comparator modules were 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 only included 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 and 45 as modules with more than 45 credits typically referred to a dissertation, and modules with fewer than 10 credits might not have enough scope for curriculum reform
- Have 10 or more students enrolled in 2021
- Have at least 2 years of pre-intervention attainment data

A total of 95 modules met the above criteria. Among these, 17 were from the Sociology course which was reformed in 2020/21, while 78 were from comparator courses 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	17	7 (41.2%)	7 (41.2%)	37.7 (21.1)
Chemistry	No	23	15 (65.2%)	12 (52.2%)	88.6 (59.3)

Criminology	No	20	9 (45.0%)	8 (40.0%)	86.3 (41.6)
Geography	No	35	13 (37.1%)	16 (45.7%)	32.3 (18.2)

3.1.2 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). The intervention intensity, according to Dr Paul Campbell's assessment, are:

- Among the 17 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 78 unreformed modules, formed a pool of comparator modules (n = 82);
- Among the remaining 13 Sociology modules, three modules ("SY1005", "SY3092" and "SY3093") that had an unknown intervention intensity score were also excluded, leaving 10 modules as reformed.

In sum, a total of 10 reformed modules remained for further analysis. Among the pool of comparator modules (n = 82), 3 were excluded as the enrolled students were exclusively international, leaving a total of 79 potential comparator modules. See **Figure 3** for the detailed module selection flow.

Overall, although the general characteristics of the reformed modules were somewhat comparable to that of the pool of comparator modules, they were not sufficiently similar (see **Table 2.2**), therefore matching is needed to identify a more robust counterfactual, i.e., a comparator group.

Table 2.2 Number of eligible modules by reformed status

	Reformed modules	Pool of comparator modules
Number of modules	10	79
Courses (n, %)	Sociology: 10 (100%)	Sociology: 4 (5.1%) Chemistry: 20 (25.3%) Criminology: 20 (25.3%) Geography: 35 (44.3%)
Compulsory modules (n, %)	5 (50%)	36 (45.6%)
Advanced modules (n, %)	4 (40.0%)	37 (46.8%)

Average number of enrolled students between 2018 to 2020 (mean, SD)	43.8 (22.4)	60.4 (38.0)
Average proportion of BAME students between 2018 to 2020 (mean, SD)	48.7 (4.6)	31.4 (15.3)
Average mark in percentile rank between 2018 to 2020 (mean, SD)	43.0 (6.6)	44.2 (8.5)

3.1.3 Module-matching procedure and results

The comparator modules were selected from the pool of eligible comparator modules. They were 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[1]
- Average percentage of BAME students from 2018 to 2020[2]
- Average attainment (percentile rank of the final module mark) among BAME students from 2018 to 2020[3]

The matching was done using the R package Matchlt [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 prior to intervention. Matching was done on a 1:1 basis, without replacement, using the nearest neighbour with no calipers. 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	Module ID	Reformed status	Propensity score
Pair 1	SY1002	Reformed	0.594

[123] For modules that only had two instead of three years of pre-intervention data, the average will be calculated for years where such data is available.

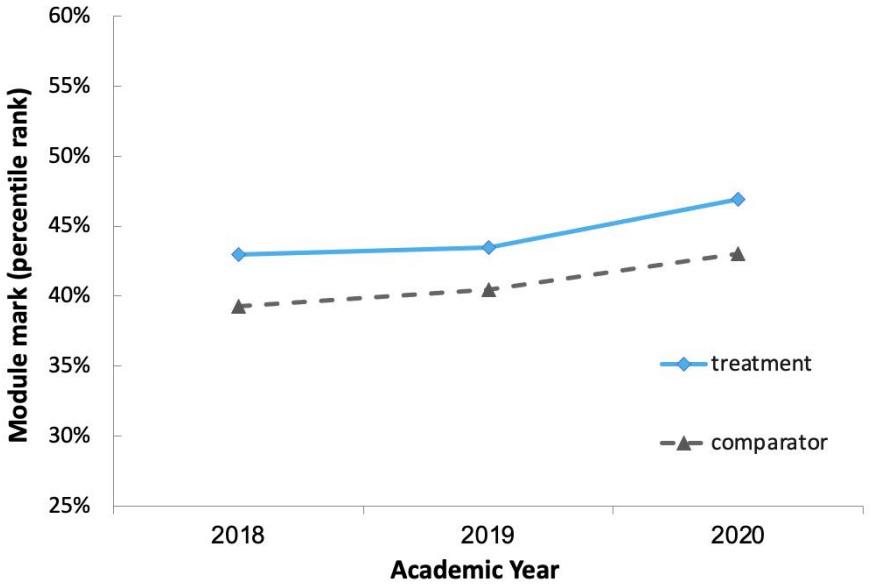
[2] Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199–236. doi: 10.1093/pan/mpl013

Pair 1	CR3028	Comparator	0.556
Pair 2	SY1004	Reformed	0.100
Pair 2	GY2416	Comparator	0.078
Pair 3	SY2008	Reformed	0.398
Pair 3	SY2078	Comparator	0.398
Pair 4	SY2089	Reformed	0.577
Pair 4	CR2026	Comparator	0.536
Pair 5	SY2091	Reformed	0.547
Pair 5	SY3095	Comparator	0.457
Pair 6	SY2094	Reformed	0.105
Pair 6	GY2431	Comparator	0.100
Pair 7	SY3079	Reformed	0.423
Pair 7	CR3023	Comparator	0.435
Pair 8	SY3090	Reformed	0.117
Pair 8	CR1004	Comparator	0.129
Pair 9	SY3094	Reformed	0.345
Pair 9	CR3030	Comparator	0.332
Pair 10	SY3098	Reformed	0.065
Pair 10	GY2432	Comparator	0.063

3.1.4 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**. On appearance, it seems that trends were parallel from 2018 to 2020. In the next section, we specify how we tested the parallel trend assumption formally.

Figure 2. Trends in weighted average module mark before intervention



3.1.5 Formal testing of the parallel trend assumption

We used a similar regression specification as the main regression (see **Section 3.6**) to test whether the pre-intervention trends of module mark (percentile rank) between treatment and comparator modules were parallel.

The regression outputs (see **Appendix 2**) showed that the trends in module mark from 2018 to 2020 of the treatment modules were not statistically different from those of the comparator modules. As a result, we think the reformed modules and the matched modules had an adequately parallel trend before the intervention.

3.2. Outcome measures

3.2.1 Definition of the outcome measure

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

Table 4. Outcome measures

Outcome measure	Data collected	Point of collection
Primary outcome: Percentile ranking of final module mark	Raw final module grades for all students of the modules of the Sociology, Criminology, Chemistry and Geography modules from academic year 2017-18 to 2020-21.	The data was routinely collected by Leicester and was provided (sent in two batches, in Aug and Nov 2021) by Leicester after the BIT-TASO data processing agreement and the Leicester-TASO data sharing agreement were signed.

	Data was anonymised before sharing.	

We used 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 is 70 or 90) will be standardised to 100 and the lowest value will be standardised to zero, making between-module difference more objective and comparable
- Percentile rank captures the difference in attainment between students rather than benchmarking against an external scale, which is better suited to the purpose of this research which focuses on the gap between White and BAME students.
- Lower risk of de-identification of module instructors

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/ upper second class honour) gaps. We acknowledge that our primary approach differs from the OfS approach, however, we think overall the benefits outweigh the risks. Nevertheless, for the output to be better comparable to other reports in this area, we also visualised the degree awarding gap using both percentile rank and percentage of students receiving either an upper second class honours or a first class in the modules, i.e. scoring 60 or higher in raw mark (see **Figure 7** and **8**).

3.2.2 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 instructors seldom give marks higher than 80 or lower than 40.

To make the results more interpretable and comparable, we also visualised the awarding gap for White and BAME students using both percentile rank and the proportion of students who were awarded upper second class honours and above (see **Section 4.4**).

3.3. Sample selection

3.3.1 Study settings

The curriculum decolonisation initiative was rolled out in 2021 among cohorts enrolled in Leicester's Sociology BA course, a full-time campus-based course.

3.3.2 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.

3.3.2.1 For modules

A total of 10 pairs (see **Table 3**) of successfully-paired modules were included for final analysis that met the criteria elaborated in the **Section 3.1.3**.

3.3.2.2 For students

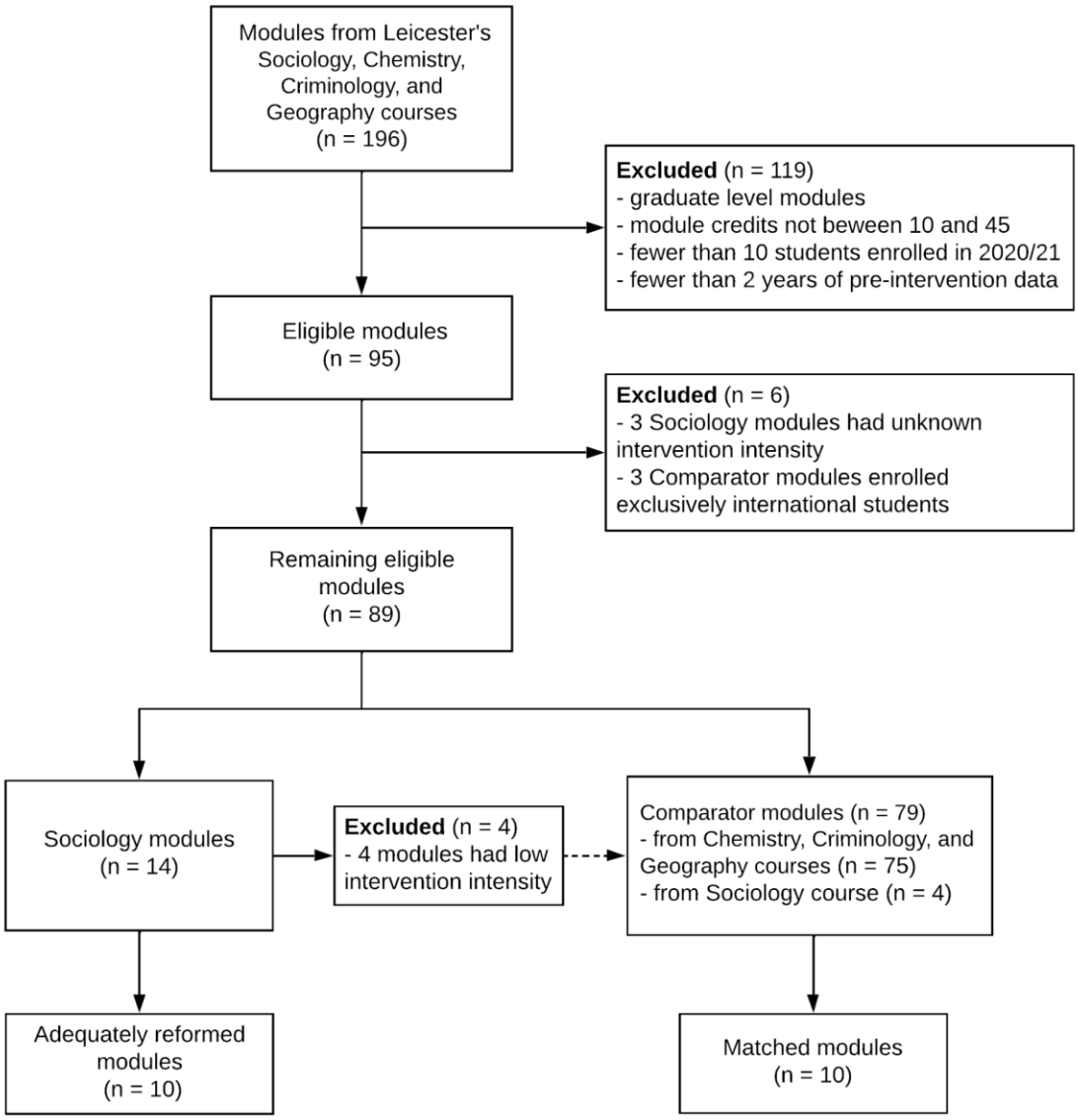
To minimise potential selection bias, within the included modules, we excluded module mark records of students whose:

- Ethnicity is unknown
- Fee payment status is other than the EU. This is because BAME students with such payment status are more likely to have been awarded scholarships to study in the UK and are not representative of general BAME students.

3.4. Module and student module marks selection flow

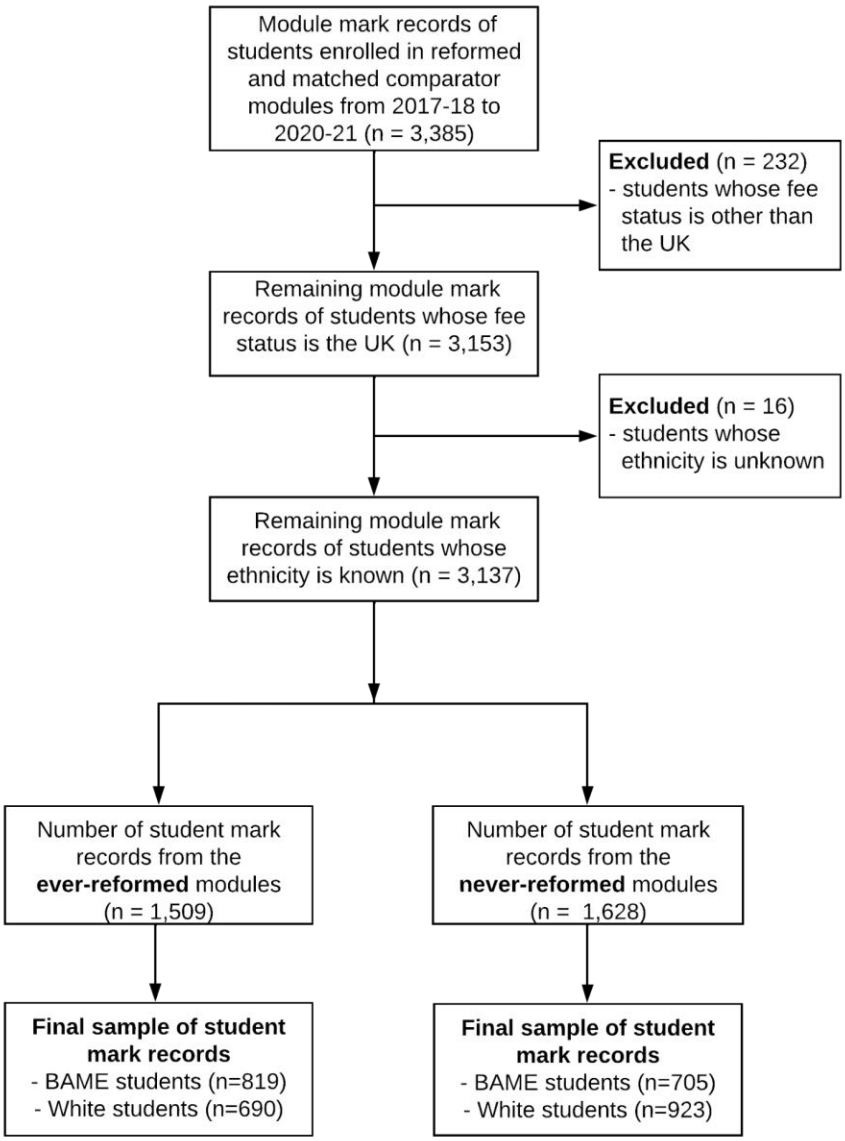
As elaborated in the **Section 3.1.2**, after applying the inclusion and exclusion criteria for module selection, a total of 10 pairs of modules were matched and retained for further analysis, see **Figure 3** for the detailed module selection process.

Figure 3. Module selection flow



Within the matched modules, we further applied the inclusion and exclusion criteria for students' module mark records (see section **3.3.2.2)** and reached a final sample (n = 2,772, out of which 1,475 were BAME students), see **Figure 4** for the module mark records selection process.

Figure 4. Student module marks selection flow



3.5. Final 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 3,137 valid observations of module mark records from 2017-18 to 2020-21 and 48.6% of them belonged to BAME students. Among the total sample, about 26.7% (838 out of 3,137) of the records took place post-intervention.

Table 5.1 Sample size of all students (including both BAME and White students)

Academic year	Reformed Modules		Comparator Modules	Overall	
	un-reformed	reformed	un-reformed	un-reformed	reformed
2017-18	438	-	425	863	-
2018-19	423	-	496	919	-
2019-20	278	-	239	517	-
2020-21	-	370	468	468	370
Total	1139	370	1628	2767	370

Table 5.2 Sample size of BAME students

Academic year	Reformed Modules		Comparator Modules	Overall	
	un-reformed	reformed	un-reformed	un-reformed	reformed
2017-18	243	-	171	414	-
2018-19	240	-	205	445	-
2019-20	134	-	122	256	-
2020-21	-	202	207	207	202
Total	617	202	705	1322	202

We also summarised how the proportion of BAME students changed over time (see **Table 5.3**). Notably, the **proportion of BAME students was higher among reformed modules than among unreformed ones in 2020/21** (54.6% vs. 44.2%). This change might be the underlying reason why we observed that average percentile rank can sometimes go up or down for both white and BAME students.[3] One potential explanation could be that modules expected to be reformed became more appealing to

[3] It may surprise some readers that the average percentile can go up for both white and BAME students. This is possible if the proportion of BAME students is not constant across years, and is an example of Yule-Simpson reversal (also known as Simpson's paradox).

BAME students thus attracting more BAME students (or fewer White students), but the legitimacy of this hypothesis is subject to the findings from implementation and process evaluation led by Leicester.

Table 5.3 Proportion of BAME students in the final sample

Academic year	Reformed Modules		Comparator Modules	Overall	
	un-reformed	reformed	un-reformed	un-reformed	reformed
2017-18	55.5%	-	40.2%	48.0%	-
2018-19	56.7%	-	41.3%	48.4%	-
2019-20	48.2%	-	51.0%	49.5%	-
2020-21	-	54.6%	44.2%	44.2%	54.6%
Total	54.2%	54.6%	43.3%	47.8%	54.6%

3.6. Analytical strategy

3.6.1 Analytical strategy

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

$$Y_{imt} = \beta_0 + \delta \ \ PostInvervention_{mt} \times EverTreated_m + \beta_1 Time_t + \beta_2 EverTreated_m + \beta_3 Gender_i + \beta_4 ModuleCompulsoryStatus_m + \beta_5 ModuleLevel_m + \epsilon_{imt}$$

Where:

- Y_i denotes the final module mark (in percentile rank) of individual i of module m in academic year t
- $\beta_0$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).
- $ModuleCompulsoryStatus_m$ is a set of dummies that denotes whether the module is compulsory or optional.
- $ModuleLevel_m$ is a set of dummies that denotes whether the module is elementary or advanced.
- ϵ_{imt} is an individual-level error term.

We use heteroskedasticity robust standard errors for all parameters.
The second analysis focuses on White students and uses the same model specification as that of the primary analysis.

The descriptive exploratory analysis focuses on the racial awarding gap between White and BAME students, and the race awarding gap results (in module mark percentile rank and % awarded upper second class and higher) are visualised using line charts.

4. Results

4.1. Description of data

Table 6 presents the baseline characteristics (averaged across the three years prior to the intervention) of the reformed versus the comparator modules. We summarised the key patterns of baseline characteristics as below:

- The proportion of advanced modules was exactly the same between reformed and unreformed modules (40%).
- The proportion of compulsory modules was similar between reformed (60%) and unreformed modules (50%).
- The average number of enrolled students, the average proportion of BAME students, and the average module mark of BAME students were all broadly similar between reformed and comparator modules.

In sum, we consider the matching quality based on base characteristics to be adequate.

Table 6. Baseline characteristics of reformed and comparator modules

	Reformed modules	Comparator modules
Number of modules	10	10
Compulsory modules (n, %)	6 (60%)	5 (50%)
Advanced modules (n, %)	4 (40.0%)	4 (40.0%)

Average number of enrolled students between 2018 to 2020 (mean, SD)	43.8 (22.4)	43.0 (29.3)
Average proportion of BAME students between 2018 to 2020 (mean, SD)	48.7 (4.59)	46.2 (18.9)
Average module mark of BAME students between 2018 to 2020 (mean, SD)	43.0 (6.62)	43.6 (8.22)

4.2. Descriptive analysis of outcomes

Table 7 presents the descriptive statistics of the primary and exploratory outcomes before and after the “Decolonising the Curriculum Toolkit" was implemented.

It is worth noting that these figures are purely descriptive, and do not imply statistical significance (see section **4.3** for results from the regression analyses). For both outcomes, we observed that, on average, BAME students’ attainment increased post-intervention in the comparator modules, while their performance decreased slightly in the reformed modules.

Table 7. Descriptive statistics of the outcomes before and after intervention

Outcome measures	Ethnicity group	Condition (reformed status)	Pre-intervention (over up to 3 years) Mean (SD)	Post-intervention Mean (SD)	Descriptive difference in difference
Module mark percentile rank	BAME students	Treatment	44.0 (29.5)	41.8 (27.9)	(41.8-44.0) - (44.4-40.7) = **-5.9 percentiles**
		Comparator	40.7 (27.2)	44.4 (30.7)	
	White students	Treatment	56.1 (27.6)	55.2 (29.6)	(55.2-56.1) - (58.0-54.1)= **-4.8 percentiles**
		Comparator	54.1 (28.0)	58.0 (28.7)	
	BAME-White gap	Treatment	12.1	13.4	(13.4-12.1) - (13.6-13.4) = **1.1 percentiles**
		Comparator	13.4	13.6	
% Achieving upper 2nd class and above	BAME students	Treatment	51.4% (50.0%)	51.0% (50.1%)	(51.0-51.4) - (56.5-42.2) = **-14.7pp**
		Comparator	42.2% (49.4%)	56.5% (49.7%)	
	White students	Treatment	70.5% (45.6%)	69.6% (46.1%)	(69.6-70.5) - (73.2-58.2) = **-15.9pp**
		Comparator	58.2% (49.4%)	73.2% (44.4%)	
	BAME-White gap	Treatment	19.1%	18.7%	(18.7-19.1) - (16.7-16.0) = **-1.1pp**
		Comparator	16.0%	16.7%	

Percentile rank. Among reformed modules, BAME students' grades were, on average, in the 44th percentile pre-intervention and the 42nd percentile post-intervention. Among the comparator modules, BAME students' grades were, on average, in the 41st pre-intervention and the 44th percentile post-intervention. In other words, we observed a relative decrease in attainment among BAME students post-intervention in treated modules compared to BAME students in comparison modules — their grades were **5.9 percentiles** lower. Similarly, we also observed a relative decrease among White students – though to a lesser extent – a decrease of **4.8 percentiles**.

Award. Among reformed modules, on average, 51.4% of BAME students were awarded upper second class honours and above pre-intervention, compared to 51.0% post-intervention. Among the comparator modules, on average, the proportion of BAME students awarded upper second class honours and above was 42.2% pre-intervention and 56.5% post-intervention. In other words, we observed a relative decrease of **14.7pp** among BAME students post-intervention in treated modules compared to those in comparison modules. We also observed a similar trend among White students – a relative decrease of **15.9pp**.

Racial awarding gap. In terms of module mark percentile rank, the racial gap widened slightly post-intervention among reformed modules and remained stable among comparator modules, representing a relative widening of **1.1 percentiles** post-intervention among treated modules. In terms of the proportion of students awarded upper second class honours and above, there was a small change in the racial awarding gap post-intervention, representing a relative narrowing of **1.1 percentiles** post-intervention among treated modules. In sum, there was limited (if any) change in racial awarding gap post-intervention among reformed modules compared to comparator modules.

4.3. Results from regression analysis

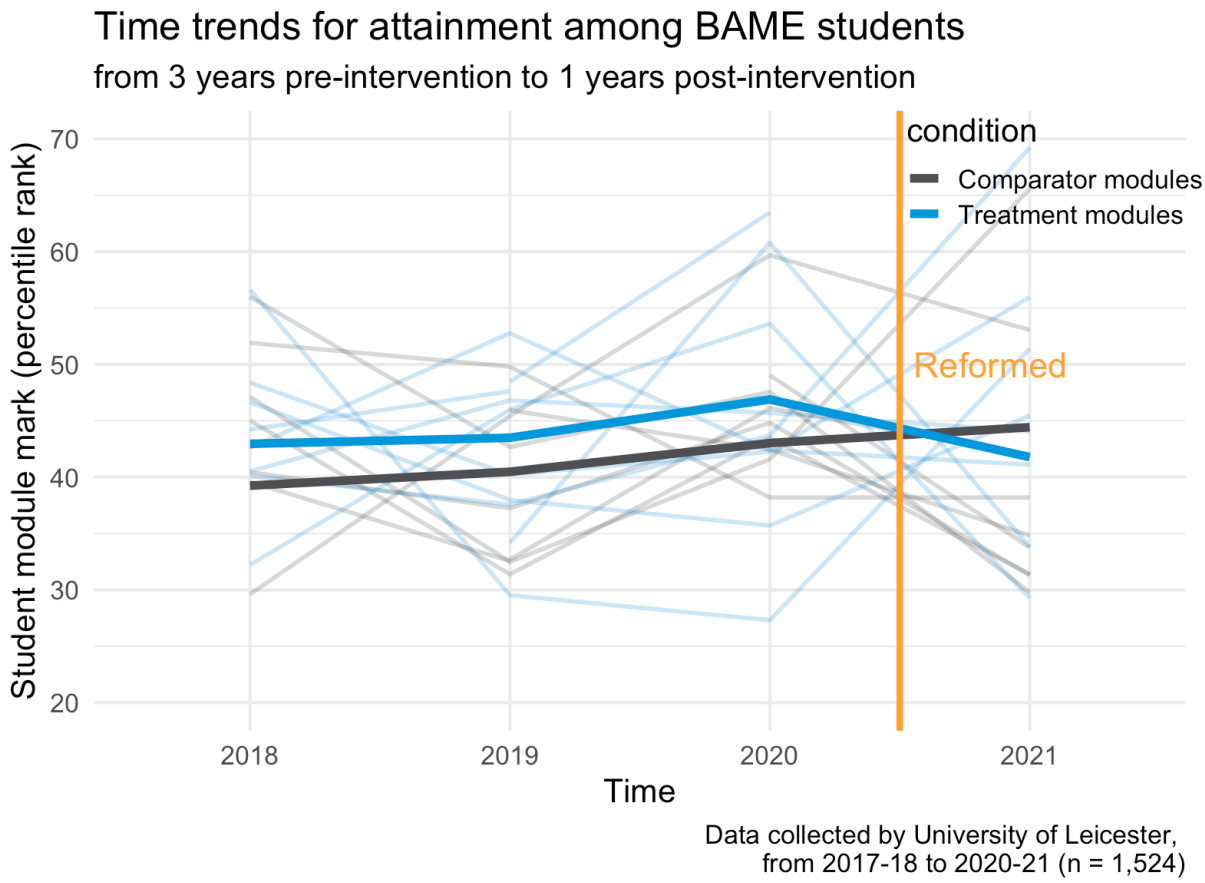
Primary analysis

There is no evidence (See **Section 3.1.4 and 3.1.5**) that suggests the parallel trends assumption was violated in any of the three years prior to intervention (see **Appendix 2** for full regression results). For this reason, we interpret the results for the primary analysis as causal.

Overall, we observed a significant negative effect on the attainment of BAME students after the 'Decolonising the Curriculum Toolkit' was implemented in the Sociology course. The estimated average treatment effect of the intervention on BAME students' attainment is **-6.63 percentiles**, **95% CI [-13.23, -0.03]**, $p = 0.05$ (see **Appendix 1** for full regression results). **Figure 5** presents the trend of attainment year by year from 2017-18 to 2020-

21. It shows that in the academic year 2021, while BAME student attainment among the comparator modules was still on a positive trajectory, it declined in the reformed modules.

Figure 5. Time trends of student attainment among BAME students



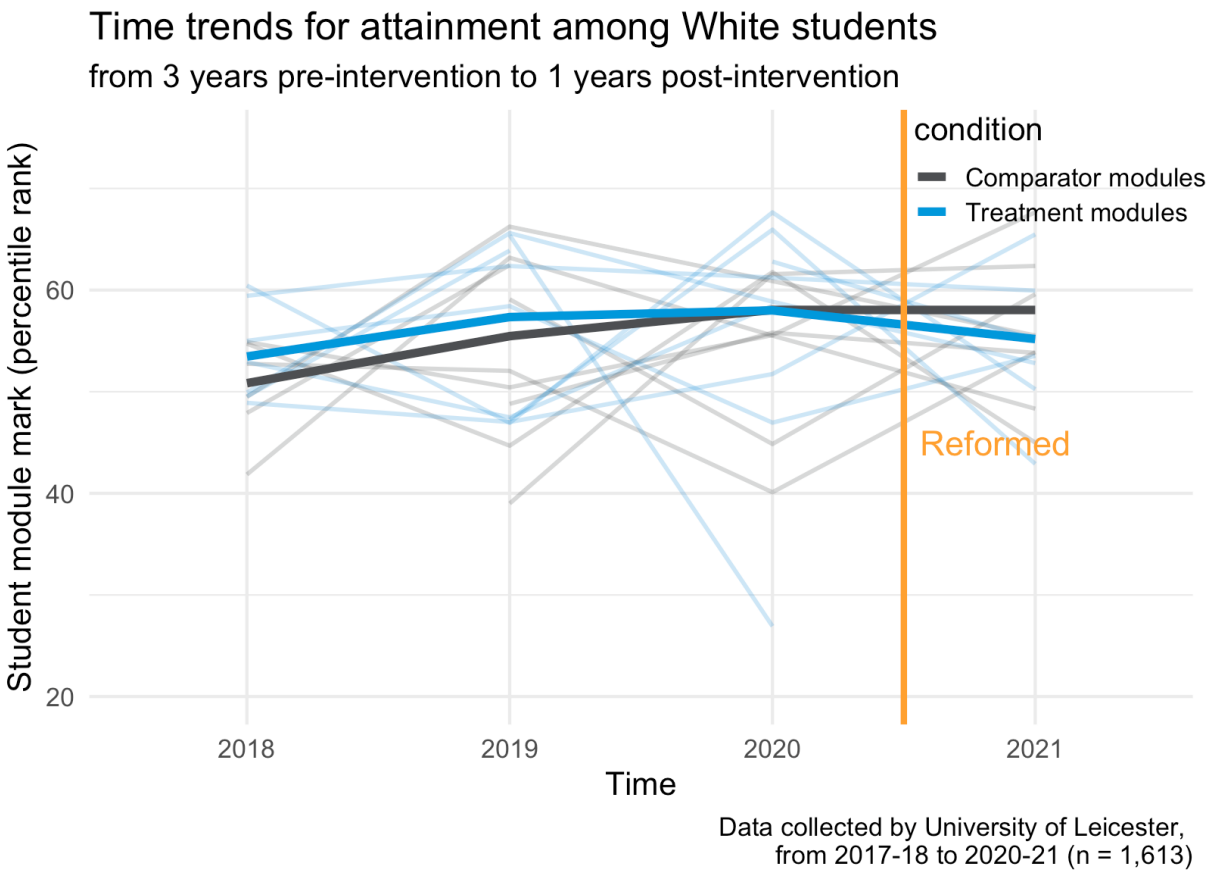
Secondary analysis

We did not check the parallel trends assumption formally for attainment among White students. By visual examination (see **Figure 6**), the trends appeared to be adequately parallel from 2018 to 2019, but they were less so from 2019 to 2020. We are therefore less confident that the results from this secondary analysis can be interpreted as causal compared to those from the primary analysis.

Overall, the attainment trends among White students were similar to those of BAME students, but the changes over time were smaller. Among comparator modules, we observed an upward trajectory both before and after the curriculum reform. Among the reformed modules, there was an upward trajectory in attainment in the years prior to the intervention and a downward trajectory after the intervention was introduced.

The estimated average treatment effect of the intervention on White students' attainment is **-3.07 percentiles**, **95% CI [-9.79, 3.64]**, *p* = 0.37 (see **Appendix 1** for full regression results).

Figure 6. Time trends of student attainment among White students



4.4. Exploratory analysis

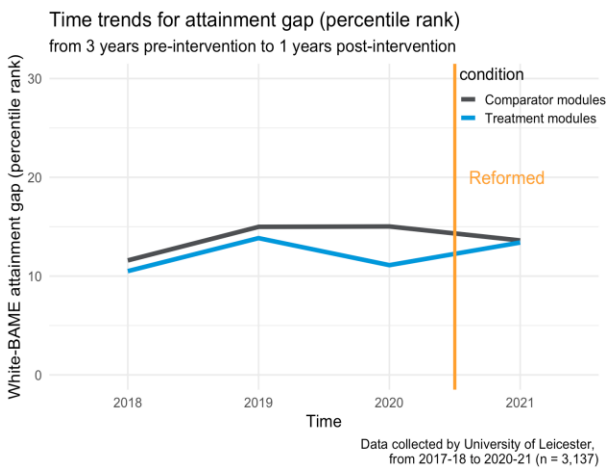
To understand the awarding gap between BAME and White students, we have presented and discussed descriptive statistics in Table 7 above. Here, to further explore this question, we have visualised the time trends of awarding gaps in terms of percentile rank (**Figure 7**) and degree awards (**Figure 8**) from 2017-18 to 2020-21.

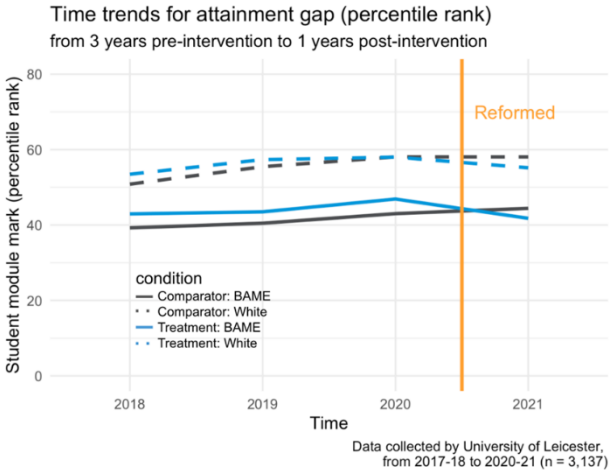
As shown in **Figure 7,** from 2018 to 2019, the awarding gap between White and BAME students was almost equal between the comparator modules (grey line) and the reformed modules (blue line). From 2019 to 2020, the awarding gap narrowed among the reformed modules but remained stable among the comparator modules.

Post intervention, i.e. in 2021, the awarding gap widened again among the reformed modules but narrowed slightly among the comparator modules. At this point, the

awarding gap was almost the same between the reformed and comparator modules.

Figure 7. Left panel: Time trends of White-BAME percentile rank gap.
Right panel: Time trends of attainment by ethnicity (percentile rank).

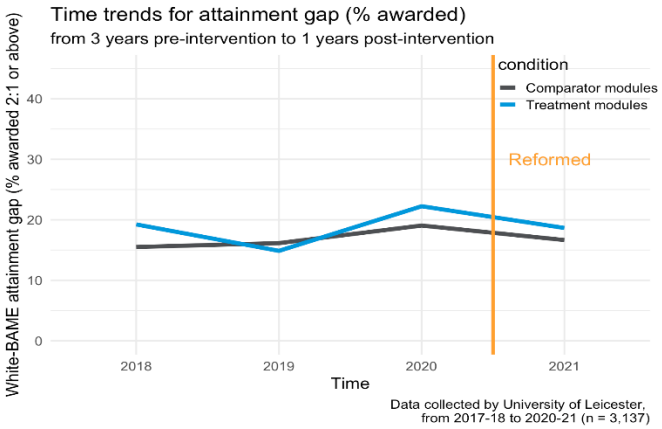


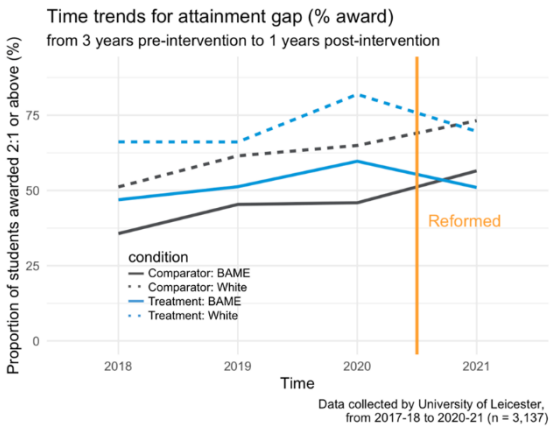


To further understand the awarding gap between BAME and White students, we also visualised the time trends of the awarding gap in terms of the proportion of students awarded upper second class honours and above (**Figure 8**) from 2017-18 to 2020-21.

The racial award gap among students in the comparator modules was on a gentle upward trajectory from 2018 to 2020, and a downward trajectory between 2020 and 2021. The racial award gap among students in the reformed modules narrowed by 5pp from 2018 to 2019, but then widened by 8pp from 2019 to 2020. However, post-intervention, the racial award gap narrowed, to a similar extent, among both reformed and comparator modules

Figure 8. Left panel: Time trends of White-BAME award gap.
Right panel: Time trends of attainment by ethnicity (% achieving award).





The exploratory analysis (as elaborated in the descriptive statistics in **Section 4.2** and shown by **figures 7** and **8**) suggests that the intervention does not seem to have had an effect on the racial awarding gap.

5. Discussion

Overall, this impact evaluation suggests that the “Decolonising the Curriculum Toolkit" might have had a negative impact on both BAME and White students’ attainment among the Sociology students at the University of Leicester. Despite these negative effects on BAME and White students’ attainment, findings from the exploratory analysis suggests that the intervention had no effect on the racial awarding gap.

The IPE should help us to understand why the intervention had a negative impact, particularly for BAME students. Two possible explanations are as follows. First, it is possible that more BAME students chose to enrol in the reformed modules — the proportion of BAME students was 10.4pp higher among reformed modules compared to unreformed ones in 2020/21 (54.6% vs. 44.2%). This self-selection then made it easier for White students to rank higher in the mark percentiles as BAME students’ attainment was on average lower than that of White students. Second, it is also possible that the timing of the curriculum reform was not ideal as it coincided with the COVID-19 pandemic. Introducing a new curriculum during the pandemic might make it more challenging for students to keep up with the course — they had to adapt to a new way of learning (remote plus in-person teaching) but had fewer available resources to help them revise for the exams (because course materials from previous years were based on different curricula).

In addition, there are two factors that might have limited the internal validity of the estimated treatment effects from our analysis. First, we did not have an objective quantification of the extent to which modules were reformed. Instead, we relied on course instructors’ self-reported data to make this assessment. It is therefore possible that the intervention intensity of some reformed modules was over- or under-estimated, and as a result our estimated treatment effects might have been over- or under-estimated. Second, there might be some spillover effects as students might simultaneously have attended both reformed and unreformed modules, which could have diluted the treatment effects. The study also has a limitation that might have constrained the generalisability of the findings. The modules that met the inclusion criteria for analysis were only a subsample of available modules as we only included 10 reformed and 10 comparator modules that met the inclusion and exclusion criteria.

In light of the above discussion, we do not recommend rolling out this toolkit (in its current form) before conducting a closer examination of how the toolkit was delivered

and perceived by teachers and students. We believe the IPE led by Leicester may shed light on what might have caused this and help contextualise these effects. If findings from the IPE suggest that there is evidence of promise in how teaching staff and BAME students received the reforms, we would recommend refining the intervention based on findings from the IPE, followed by further impact evaluation, with a larger sample and over a longer time period.

Appendix 1 — primary analysis (1) and secondary analysis (2) regression outputs

	BAME students	White students
(Intercept)	43.88 **	59.51 **
	CI [39.04, 48.72], p < 0.001	CI [54.10, 64.93], p <0.001
Post-intervention: Yes ×	-6.63 *	-3.07
Reformed: Yes	CI [-13.23, -0.03], p = 0.05	CI [-9.79, 3.64], p = 0.37
Post-intervention: Yes	1.09	0.02
	CI [-4.71, 6.89], p = 0.71	CI [-5.19, 5.24], p = 0.99
Reformed: Yes	3.56 *	0.31
	CI [0.20, 6.91], p = 0.04	CI [-3.07, 3.68], p = 0.86
Academic Year: 2018	-4.13 +	-5.34 *
	CI [-8.84, 0.59], p = 0.09	CI [-9.62, -1.07], p = 0.01
Academic Year: 2019	-3.12	-1.18
	CI [-7.65, 1.40], p = 0.18	CI [-5.51, 3.14], p = 0.59
Gender: Male	-6.86 **	-6.68
	CI [-10.63, -3.09], p < 0.001	CI [-9.83, -3.54], p < 0.001
Module level: Advanced	-1.43	1.33
	CI [-5.04, 2.18], p = 0.44	CI [-2.69, 5.35], p = 0.52
Compulsory status: Compulsory	1.84	-0.57
	CI [-1.54, 5.21], p = 0.29	CI [-4.42, 3.28], p = 0.77
N	1524	1613
R2	0.01	0.02

Standard errors are heteroskedasticity robust. ** p < 0.01; * p < 0.05; + p < 0.1.

Appendix 2 — Formal testing of parallel trends assumption among BAME students

	Parallel trends assumption check
(Intercept)	43.78 **
	CI [38.01, 49.55], p < 0.001
Gender: Male	-6.90 **
	CI [-10.68, -3.12], p < 0.001
Compulsory status: Compulsory	1.82
	CI [-1.57, 5.20], p = 0.29
Module level: Advanced	-1.45
	CI [-5.07, 2.17], p = 0.43
Academic Year: 2018	-4.42
	CI [-11.03, 2.19], p = 0.19
Academic Year: 2019	-2.57
	CI [-9.01, 3.88], p = 0.43
Academic Year: 2021	1.22
	CI [-5.44, 7.88], p = 0.72
Reformed: Yes	3.80
	CI [-3.68, 11.28], p = 0.32
Academic Year: 2018 × Reformed: Yes	0.48
	CI [-8.74, 9.70], p = 0.92
Academic Year: 2019 × Reformed: Yes	-1.04
	CI [-10.15, 8.07], p = 0.82
Academic Year: 2021 × Reformed: Yes	-6.88
	CI [-16.25, 2.50], p = 0.15
N	1524
R2	0.01

Standard errors are heteroskedasticity robust. ** p < 0.01; * p < 0.05; + p < 0.1.