Analysis report

University of Kent Impact Evaluation Analysis

November 2022

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

QA: James Lawrence (Head of Quantitative Research, BIT)

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

The study was pre-registered on Open Science Framework

1. Summary

Background

The Centre for Transforming Access and Student Outcomes in Higher Education (henceforth TASO) has funded the University of Kent (henceforth Kent) and commissioned the Behavioural Insights Team (henceforth BIT) to evaluate the impact of their "Diversity Mark" programme (an initiative that seeks to diversify the current Eurocentric curriculum) on reducing awarding gaps between Black, Asian and minority ethnic (BAME) students and White students.

Aims

To evaluate whether and to which extent Kent's 'Diversity Mark' initiative reduced the awarding gaps between BAME and White students.

Intervention

The "Diversity Mark" initiative is a collaborative response to Kent students' call for more diverse curricula. The School of Sociology, Social Policy and Social Research (henceforth SSPSSR), students, and library services worked together to audit 19 core undergraduate modules offered in the two campuses and explored ways to incorporate BAME authors and perspectives into those modules. The initiative was first piloted in 2018-19, and another module was piloted in 2020-21.

Design

The study is a matched difference-in-differences with repeated cross-sections. The analysis compares students' attainment trend among the modules that implemented the Diversity Mark Initiative (treatment modules) with similar comparator modules that didn't implement the initiative.

Outcome measures

The primary outcome measure is a student's 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 2018-19 between reformed vs. matched unreformed modules. It focuses on BAME students only. The secondary analysis repeats the primary analysis for White students. Additional descriptive charts are made to illustrate the change in awarding gaps of reformed vs. comparator modules before and after the Diversity Mark Initiative.

Results

Among the modules matched for analyses (4 reformed modules, 4 comparator modules), we did not observe a significant effect of the Diversity Mark Initiative on improving attainment in terms of module mark percentile rank among the BAME students — the average difference in attainment between reformed and unreformed modules post-intervention versus pre-intervention was not statistically significant (2.0 percentile rank, 95% CI [-2.20, 6.21]), $p = 0.35$. We observed a marginally positive difference among the White students (3.45 percentile rank, 95% CI [-0.13, 7.03]), $p = 0.06$.

Conclusions

Overall, we didn't find conclusive evidence supporting the effectiveness of the Diversity Mark Initiative in reducing the racial awarding gap among the SSPSSR students at the University of Kent. However, we also didn't find evidence that suggests it might backfire: the observed trend among BAME students' attainment before and after the initiative, though not significant, was positive. Therefore, we consider the initiative an innovative approach to address the racial awarding gap that is worth further testing.

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 Kent 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 Kent. Kent's colleagues shared background information of the Diversity Mark Initiative and their preliminary project report with BIT as well as sharing GDPR-compliant individual-level module data. In addition, they also helped 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	James Lawrence	Evaluation Supervisor and quality assurance
TASO	Sarah Chappell	Project liaison
TASO	Dr Helen Lawson	Research/project lead
KENT	Professor Kathleen M Quinlan	Partner lead
KENT	Dr Barbara Adewumi	Partner co-investigator
KENT	Dr Ellen Dowie	Partner co-investigator
KENT	Dr Miyoung Ahn	Research associate

2.2. Aims

Some researchers have argued that the race awarding gap could be attributed to the 'whiteness' of the curriculum (Mountford-Zimdars et al., 2015; Mcduff, Tatam, Beacock, & Ross, 2018). According to the BME Student Voice Project in 2016, Kent currently has

Eurocentric curricula and lacks BME authors. Kent's own students were aware of the lack of diversity and have voiced their desire toward more diverse curricula (e.g., Kent Union's "Diversify my Curricula" campaign and the Decolonise UoK campaign).

However, little empirical research has been done to investigate the causal relationship between diversifying curricula and the race awarding gap. In light of this background, this study aims to offer initial evidence of potential benefits of diversifying curricula by evaluating whether and to which extent Kent's 'Diversity Mark' initiative reduced the awarding gaps between BAME and white students.

2.2.1 Research questions

The lack of empirical evidence led us to ask the following research questions.

The primary research question:
How did Kent's 'Diversity Mark' initiative affect the attainment of BAME students?

The secondary research question:
How did Kent's 'Diversity Mark' initiative affect the attainment of White students?

The exploratory research question:
How did Kent's 'Diversity Mark' initiative affect the awarding gap between White and BAME students?

2.2.2 Research hypotheses

We hypothesize that undergraduate SSPSSR core modules that have diversified their curricula will have smaller post-intervention White/BAME awarding gaps than comparator modules that did not diversify their curricula.

2.2.3 Rationale for choosing comparators

Comparator modules were chosen to establish plausible counterfactuals, for participation in the 'Diversity Mark' initiative was voluntary for module instructors, therefore module reformation could not be (nor could be considered) randomly assigned. See Section 5.1 for details on matching methodology.

2.3. Intervention

2.3.1 Overview of the Diversity Mark Initiative

University of Kent's 'Diversity Mark' initiative is a collaboration between the Student Success Team, students, and library services designed to include more BAME perspectives in the formal curriculum, making it more culturally sensitive. The intervention is based on the assumption that a diversified curriculum will affect students'

interest and interactions with their teachers on academic matters relating to their programme of study.

The initiative involved training students to conduct a reading list audit, run student focus groups, work with the library to identify potential further resources, and present their findings back to module convenors. These activities culminated in an individual interview with module convenors conducted by the intervention lead. In the interview, module convenors were asked to reflect on the role of reading lists in relation to the rest of their teaching, to consider challenges of diversifying their curriculum, to outline how diversity currently features in their teaching, and to describe their plans for changing their curriculum to make it more culturally sensitive.

The intervention was piloted across first year modules in the School of Social Policy, Sociology, and Social Research (SSPSSR) in the academic year 2018-19. Modules were deemed reformed if module convenors demonstrated commitment to creating more culturally sensitive curricula in various ways. These included:

- Detailing plans for diversifying reading lists,
- Giving examples of how they would integrate diversity across the module rather than confining it to a single topic or week,
- Showing awareness of the impact of negative portrayals of racially or ethnically minoritised people,
- Evidencing critical awareness of race and ethnicity as it pertained to their curriculum.

2.3.2 Implementation of the Diversity Mark Initiative

The Diversity Mark Initiative was first piloted at SSPSSR during the Summer of 2018. The process involves three main stages:

- **Auditing.** The 2017-18 reading list of core undergraduate modules offered at SSPSSR at two campuses. A total of 19 core modules' reading lists were audited. All Stage 1 modules across all SSPSSR on Campus 1 programmes were examined, which included three degree programmes: Social Sciences, Criminal Justice and Social Work.
- **Feedback.** After the auditing, the module instructors received feedback on how to diversify their reading lists. Module instructors were sent feedback to their reading lists, along with a survey with five open-ended questions:
    1. What is your understanding of the function and purpose of the reading list in relation to the rest of teaching?

2. To what extent do you think there are challenges to the development of a more inclusive curriculum in your subject area? Please outline some of these perceived challenges.
3. Please outline some of the ways that questions of diversity and demographic difference currently feature in your teaching.
4. Are there any plans to change?
5.Do you have any wider thoughts on tackling attainment gaps and diversifying the curriculum?

- **Reform.** Every module instructor (except one) responded to the survey. Following the module audit and survey, five module instructors of audited modules indicated intentions to change their curriculum and plans to incorporate more BAME authors and perspectives into their modules based on the feedback they received from the audit. Those five modules' curricula were reformed after the academic year 2017-18. Another module instructor showed interest and his module was reformed in 2020-21 (see **Table 2** in Section 3.2 for details).

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 didn't reform their curricula.

3.1.1 Module matching criteria

The comparator modules were chosen from a pool of unreformed modules based on how similar they were to the reformed modules pre-intervention in the following characteristics:

- Campus (campus 1 or campus 2)
- Whether the module is for stage 1 or not (all reformed modules were stage 1 modules)
- Whether the module is textbook-driven (only non-textbook-driven modules have enough scope for curricula diversification)
- Whether the module has at least one year of pre-intervention data available
- Average number of enrolled students from t-4 to t-1, where t is the first year that the reformed curricula were taught (t = 2018-19 for modules 3,4,5; t = 2020-21 for module 20)

- Average percentage of BAME students from t-4 to t-1, where t is the first year that the reformed curricula were taught
- Average attainment (percentile rank of the final module mark) among BAME students from t-4 to t-1, where t is the first year that the reformed curricula were taught

The reformed status and key module characteristics are presented in **Table 2**.

Table 2. Key characteristics for reformed and candidate comparator modules

moduleID	Campus	Stage	Reformed	Availability of 3 years of pre-intervention data	Textbook driven	Included in matching
Module 01	Campus 1	1	Yes (in 18/19)	No	No	No
Module 02	Campus 1	1	Yes (in 18/19)	No	No	No
Module 03	Campus 1	1	Yes (in 18/19)	Yes	No	Yes
Module 04	Campus 2	1	Yes (in 18/19)	Yes	No	Yes
Module 05	Campus 2	1	Yes (in 18/19)	Yes	No	Yes
Module 06	Campus 1	1	No	Yes	No	Yes
Module 07	Campus 1	1	No	No	Yes	No
Module 08	Campus 1	1	No	No	Yes	No
Module 09	Campus 1	1	No	No	No	No
Module 10	Campus 1	1	No	Yes	No	Yes
Module 11	Campus 1	1	No	Yes	No	Yes
Module 12	Campus 1	3	No	No	NA	No
Module 13	Campus 1	1	No	Yes	No	Yes
Module 14	Campus 1	2	No	No	NA	No
Module 15	Campus 1	2	No	No	NA	No
Module 16	Campus 1	1	No	Yes	Yes	Yes
Module 17	Campus 2	1	No	Yes	No	Yes
Module 18	Campus 2	1	No	Yes	No	Yes
Module 19	Campus 2	1	No	Yes	No	Yes
Module 20	Campus 2	1	Yes (in 20/21)	Yes	No	Yes
Module 21	Campus 2	1	No	Yes	No	Yes
Module 22	Campus 2	1	No	Yes	No	Yes
Module 23	Campus 2	1	No	Yes	No	Yes

3.1.2 Module inclusion and exclusion criteria

Modules were excluded from further analysis for the following reasons:

- **Module 12, 14, and 15** were excluded as they are not stage 1 modules.
- **Module 1, 2, 7, 8, 9** were excluded due to poor data availability of pre-intervention years (having no or only 1 year of pre-intervention data).
- **Module 16** was excluded as it's textbook-driven, in which case the scope for diversifying curricula is severely constrained.

After excluding the above modules, a total of 14 modules remained: 4 of them were reformed and 10 of them were candidate comparators (summarised in the final column of **Table 2**). Some student records were also excluded (see **Section 3.3.2** sample inclusion and exclusion criteria).

3.1.3 Module-matching procedure and results

The matching was done using the R package Matchlt [1]. Each Module was matched based on the following characteristics:

- Campus
- Average number of enrolled students from t-4 to t-1
- Average percentage of BAME students from t-4 to t-1
- Average module attainment (percentile rank of the final module mark) among BAME students from t-4 to t-1

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 module 3, 4, 5 (reformed in 2018-19) and module 20 (reformed in 2020-21). **Table 3** presents the propensity scores of the reformed modules pairing with four comparator modules that had the closest propensity scores.

Table 3. Propensity scores of reformed vs. comparator modules

Reformed module	Propensity score	Comparator module	Propensity score
Module 3	0.3474	Module 13	0.4138
Module 4	0.6668	Module 18	0.6203

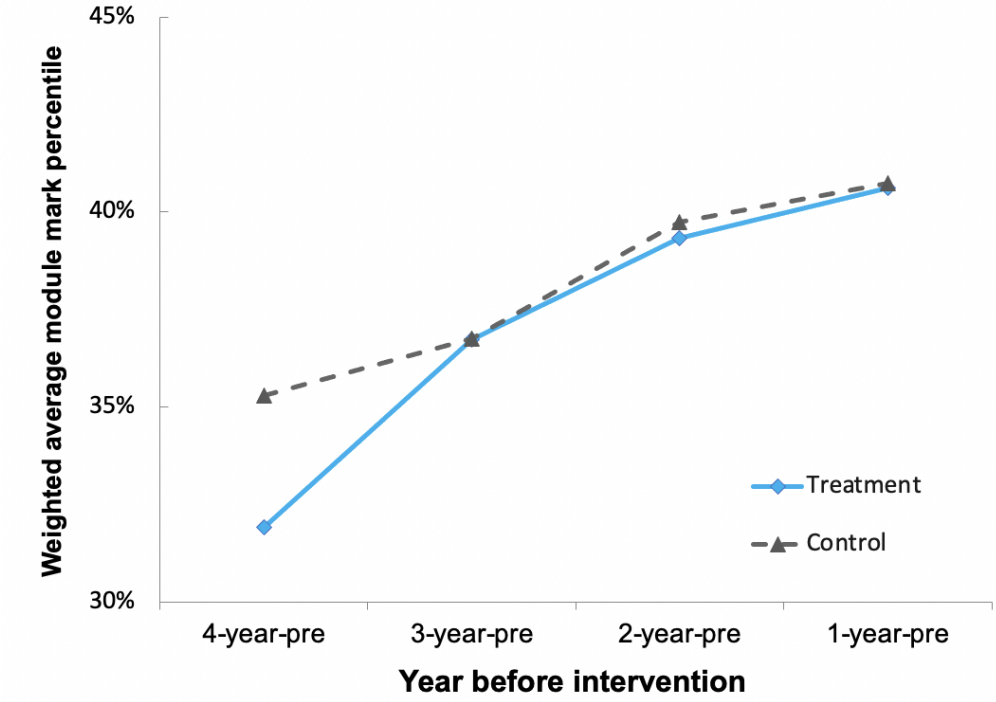
[1] Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199–236. doi: 10.1093/pan/mpl013

Module 5	0.3911	Module 21	0.4107
Module 20	4.64e-11	Module 17[2]	1.00e+00

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 4 years prior to intervention. We then plotted the parallel trends in Figure 2. It appears that the trends were parallel up to 3 years prior to intervention. In the next section, we specify how we test the parallel trend assumption formally.

Figure 2. Trends in weighted average[3] module mark before the intervention



3.1.5 Formal testing of the parallel trend assumption

We used a similar regression specification as the main regression (see **Section 11**) to

[2] The propensity score is 0 for module 20 and 1 for module 17 because there was only one module that was treated in 2020-21, therefore the logistic regression fit perfectly. Despite the sharp disparity, the other variables match reasonably well so we consider module 17 as an adequate match for module 20.
[3] Since modules vary greatly in number of enrolled BAME students, we weighted the module attainment by BAME students count.

test whether the pre-intervention trends between treatment and comparator modules were parallel.[4] The regression outputs (using data up to 2019-20) showed that although the trend in the 4th pre-intervention year appeared non-parallel in the plot, it was not statistically significant compared to the trend during the first pre-intervention year. As a result, the reformed modules and the matched modules had an adequately parallel trend before the intervention for us to interpret this relationship as causal.

3.2. Outcome measures

3.2.1 Definition of the outcome measure

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

Table 4. Outcome measures

Outcome measure	Data 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 from academic year 2014-15 to 2020-21. Data was anonymised before sharing.	The data is routinely collected by Kent and was provided (sent in two batches, in Aug and Nov 2021) by Kent once the BIT-TASO data processing agreement and the KENT-TASO data sharing agreement were signed.

We use percentile rank of module mark as our primary 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 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 (See Section 12.2 of the Trial Protocol for details)

[4] In the formal testing we interact each pre-intervention year with the treatment dummy to identify whether treated and control modules have different trends in each year pre-intervention. In the main analysis we interact with a more general 'post' dummy with the treatment dummy to increase power.

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 upper second class and above 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 visualised the degree awarding gap using both percentile rank and percentage of students awarded upper second class and above in the modules (see **Figure 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. Descriptive analysis (using data up to 2019-20) shows that the mean raw mark was 59.6 for White students and 55.4 for BAME students, whereas the mean percentile rank was 53.3 for White students and 38.9 for BAME students. Thus the awarding gap might seem wider when using percentile rank, however this metric is useful in establishing students' standing in relation to others. As a result, we visualised the awarding gap for White and BAME students using both percentile rank and the percentage of students who achieved upper second class and above (defined as students achieving a raw score of 59.5 or above).

3.3. Sample selection

3.3.1 Study settings

The Diversity Mark Initiative was piloted among cohorts enrolled in Kent's SSPSSR degree courses that took place in two campuses. While the named degree courses are similar and follow the same assessment patterns and overall School and University policies, the two campus cohorts are separate. At one campus, 147 undergraduates were enrolled in SSPSSR degree courses in 2017-18; and on the other 337 were enrolled.

3.3.2 Inclusion and exclusion criteria

The sample comprises BAME and White students'[5] final module marks of Stage 1 core social science modules in the following academic years: 2014-15, 2015-16, 2016-17, 2017-18, 2018-19, 2019-20, and 2020-21.

[5] We used a binary classification of ethnicity: White British and White Irish were counted as 'white' and anything else except "unknown" was counted as BAME; "unknown" were excluded from further analysis.

3.3.2.1 For modules

All modules (see **Table 2**) were worth 15 credits and taught over 12 weeks by an SSPSSR staff member in one of the degree programmes in SSPSSR. In addition, as specified in **Section 5.1 and 5.2**, all included modules must be:

- Stage 1 core modules
- Having at least three years of pre-intervention administrative data
- Fitting a typical reading list pattern (i.e. not textbook-driven), following a standard assessment pattern which typically consisted of essay assignments, short research projects, presentations and an end of module exam.

3.3.2.2 For students

To minimise potential bias, within the included modules, we excluded students whose:

- Ethnicity is unknown
- Students whose fee payment status is other than the UK, i.e., non-local students (this is consistent with the approach of the OfS). We excluded international students as BAME students from abroad often come with scholarships, and their attainment may not be representative of that of the typical BAME students.

3.4. Module and student module marks selection flow

As elaborated in the **section 3.1.2**, Kent provided module-level and individual-level data for a total of 23 SSPSSR modules, out of which 14 were eligible and a total of 8 modules were matched and retained for further analysis, see **Figure 3** for the detailed module selection process.

Within the matched modules, we further applied the inclusion and exclusion criteria for students' module mark records as specified (see section **3.3.2.2**). In the end, we are left with the final sample for analysis (n = 6,854, out of which 2,667 were BAME students), see **Figure 4** for the module mark records selection process.

Figure 3. Module selection flow

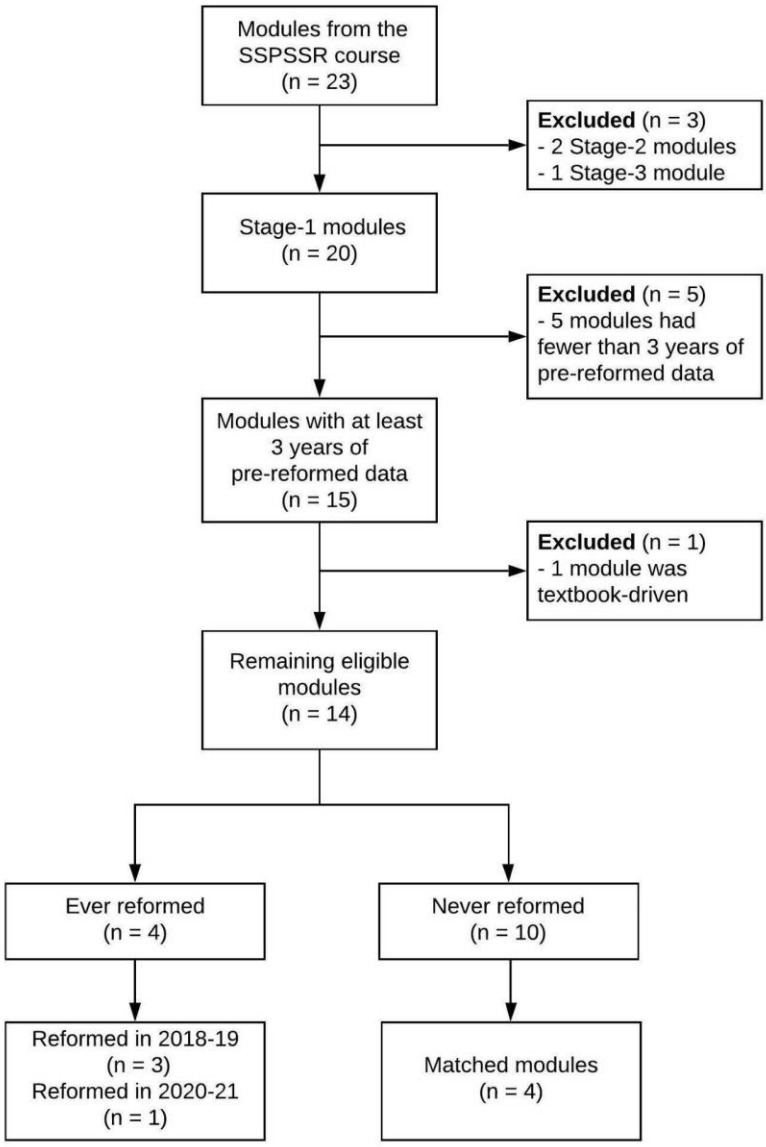
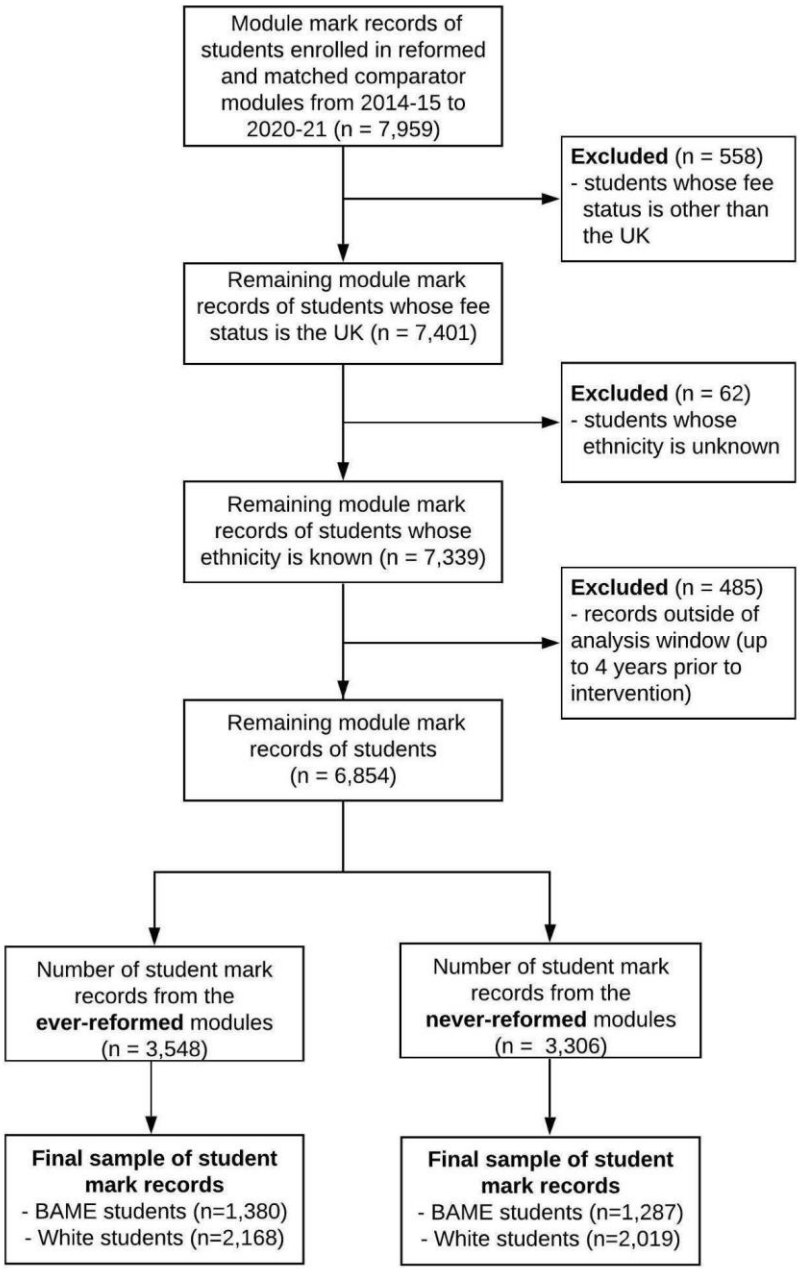


Figure 4. Student module marks selection flow



3.5. Final sample size

On average, 38.9% of all module mark records belonged to BAME students and 61.1% to White students. And as shown in **Table 5.1**, among all students (White and BAME), about 19.3% (1,320 out of 6,854) of the records took place post-intervention. Among the BAME students (see **Table 5.2**), roughly 20.7% (551 out of 2,667) of the records took place post-intervention.

The final sample size was largely in line with the sample size that was estimated in the trial protocol (n = 7,249, out of which 2,813 were BAME students), which means the power analysis conclusion from the trial protocol — we are powered to detect a difference of **6.50 percentile rank** for the primary analysis — is largely applicable for our final analysis.

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

Academic year	Module 3,4,5		Module 20		Comparator Modules	Overall	
	un-reformed	reformed	un-reformed	reformed	un-reformed	un-reformed	reformed
2014-15	296	-	-	-	180	476	-
2015-16	343	-	-	-	335	678	-
2016-17	356	-	173	-	531	1060	-
2017-18	432	-	222	-	652	1306	-
2018-19	-	388	197	-	533	730	388
2019-20	-	400	209	-	564	773	400
2020-21	-	339	-	193	511	511	532
Total	**1427**	**1127**	**801**	**193**	**3306**	**5534**	**1320**

Table 5.2 Total sample size for BAME students

Academic year	Module 3,4,5		Module 20		Comparator Modules	Overall	
	un-reformed	reformed	un-reformed	reformed	un-reformed	un-reformed	reformed
2014-15	127	-	-	-	79	206	-
2015-16	145	-	-	-	127	272	-
2016-17	118	-	52	-	169	339	-
2017-18	155	-	76	-	238	469	-
2018-19	-	175	88	-	251	339	175
2019-20	-	143	68	-	201	269	143
2020-21	-	159	0	74	222	222	233
Total	**545**	**477**	**284**	**74**	**1287**	**2116**	**551**

We also summarised how the proportion of BAME students changed over time (see **Table 5.3**). Overall, the proportion of BAME students was slightly higher after reform (41.7% vs. 38.2%) . This change might be the underlying reason why we observed that average percentile rank can sometimes go up for both white and BAME students.[6]

Table 5.3 Proportion of BAME students in the final sample

Academic year	Module 3,4,5		Module 20		Comparator Modules	Overall	
	un-reformed	reformed	un-reformed	reformed	un-reformed	un-reformed	reformed
2014-15	42.9%	-	-	-	43.9%	43.3%	-
2015-16	42.3%	-	-	-	37.9%	40.1%	-
2016-17	33.1%	-	30.1%	-	31.8%	32.0%	-
2017-18	35.9%	-	34.2%	-	36.5%	35.9%	-
2018-19	-	45.1%	44.7%	-	47.1%	46.4%	45.1%
2019-20	-	35.8%	32.5%	-	35.6%	34.8%	35.8%
2020-21	-	46.9%	-	38.3%	43.4%	43.4%	43.8%
Total	38.2%	42.3%	35.5%	38.3%	38.9%	38.2%	41.7%

[6] 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).

3.6. Analytical strategy

The primary analysis focuses on BAME students only, including data from academic year 2014-15 to 2020-21. The analysis is a difference-in-difference regression with multiple pre-intervention and post-intervention data points, and the intervention took place in different years. The OLS regression model is specified as follows:

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

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 values from 2014-15 to 2020-21.
- $MatchedPair_m$ is a set of dummies that denotes one of four pairs of modules matched by propensity scores based on module characteristics.
- $Gender_i$ denotes the gender of participant i gender (0 = female; 1 = male)
- $Campus_m$indicate whether module m was taught (0 = campus 1; 1 = campus 2)
- 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 attainment 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 of the 4 pairs of matched modules (averaged across the four years prior to the intervention). We summarised the key patterns of baseline characteristics as below:

- The average proportion of BAME students was broadly similar across modules and pairs, ranging from 36% to 40%.
- Except for the first pair of modules, all modules had an average of more than 100 students enrolled during the 4 pre-intervention years.
- Although the racial gap in attainment varies by pair, it was generally sizable. White students on average scored 8 ~16 percentile rank higher than BAME students. In addition, White students were also much more likely to have achieved an upper second class and above than BAME students, and in some extreme cases (e.g. Module 20), almost twice as many White students were awarded upper second class and above compared to BAME students.

Table 6 Baseline (from 1 to 4 years pre-intervention) characteristics of matched modules for final analysis

Pair name	Module ID	Reformed status	Year reformed	Average no. of enrolled students Mean (SD)	Average % of BAME students Mean (SD)	Average percentile rank of module mark Mean (SD)		Average % of students achieving 2nd class or above Mean (SD)	
						BAME	White	BAME	White
Pair 1	Module 3	Yes	2018-19	39 (6)	39% (15%)	45.5 (19.7)	53.0 (12.1)	42.4% (27.0%)	55.1% (17.7%)
Pair 1	Module 13	No	NA	40 (6)	40% (14%)	42.1 (7.8)	54.0 (5.3)	46.1% (16.8%)	63.8% (7.8%)
Pair 2	Module 4	Yes	2018-19	176 (45)	38% (5%)	37.5 (1.3)	52.9 (6.2)	40.5% (3.2%)	67.4% (8.4%)
Pair 2	Module 18	No	NA	153 (32)	38% (5%)	38.3 (3.0)	52.6 (6.1)	42.3% (7.3%)	66.3% (8.0%)
Pair 3	Module 5	Yes	2018-19	143 (16)	39% (16%)	37.8 (8.0)	50.7 (5.9)	39.0% (16.1%)	57.7% (9.1%)
Pair 3	Module 21	No	NA	151 (101)	37% (3%)	38.8 (5.1)	49.0 (14.3)	48.3% (18.0%)	53.4% (25.0%)
Pair 4	Module 20	Yes	2020-21	173 (38)	36% (8%)	35.9 (2.6)	52.0 (3.7)	25.6% (7.5%)	51.6% (5.5%)
Pair 4	Module 17	No	NA	129 (42)	36% (10%)	36.4 (6.9)	49.0 (7.4)	26.7% (12.0%)	41.8% (16.1%)

4.2. Descriptive analysis of outcomes

Table 7 presents the descriptive statistics of the primary and exploratory outcomes before and after the Diversity Mark Initiative was implemented. Among reformed modules, BAME students on average scored 40.6 percentile rank post-intervention, compared to 37.4 percentile rank prior to intervention, representing an increase of 3.2 percentile. Among the comparator modules, the increase was only 0.5 percentile rank (from 38.9 to 39.4). In other words, we observed a small increase (**2.7** percentile rank) in attainment among BAME students post-intervention. Similarly, we observed a small increase (3.3 percentile rank) among White students.

As for the proportion of students awarded an upper second class and above, we also observed a minor increase of **2.5pp** among BAME students. Among the reformed modules, only 34.9% of BAME students were awarded an upper second class and above pre-intervention, whereas 42.8% achieved so post-intervention, representing an increase of **7.9pp**. Among the unreformed modules, the increase was only **5.4pp** within the same time horizon. However, we didn't observe a similarly positive trend among White students (**-0.4pp**).

It is worth noting that these figures were purely descriptive, and do not imply statistical significance. See section **4.3** for results from the regression analyses.

Table 7 Descriptive statistics of the outcomes before and after intervention

Ethnicity group	Outcome measures	Condition (reformed status)	Pre-intervention (average over 4 years)[7] Mean (SD)	Post-intervention (average over 1~3 years)[8] Mean (SD)	Descriptive difference in difference
BAME students	Module mark percentile rank	Comparator	38.9 (27.5)	39.4 (27.0)	(40.6-37.4) - (39.4-38.9) = **+ 2.7**
		Reformed	37.4 (26.5)	40.6 (28.1)	
	% Achieving 2nd class and above	Comparator	38.8% (47.8%)	44.2% (49.7%)	(42.8-34.9) - (44.2-38.8) = **+ 2.5pp**
		Reformed	34.9% (47.7%)	42.8% (49.5%)	
White students	Module mark percentile rank	Comparator	53.5 (28.5)	53.7 (27.5)	(55.2-51.7) - (53.7-53.5) = **+ 3.3**
		Reformed	51.7 (27.4)	55.2 (29.1)	
		Comparator	59.5% (49.1%)	65.3% (47.6%)	(62.7-57.3) - (65.3-59.5) = **-**

[7] For module 3, 4 , 5, 13, 18, and 21, we calculated the average between the academic year 2014-15 to 2017-18; for module 20 and 17, we calculated the average between 2016-17 to 2019-20.
[8] For module 3, 4 , 5, 13, 18, and 21, the post-intervention years started from the academic year 2018-19 to 2020-19; for module 20 and 17, the only post-intervention year was 2020-21.

	% Achieving upper second class and above	Reformed	57.3% (49.5%)	62.7% (48.4%)	**0.4pp**

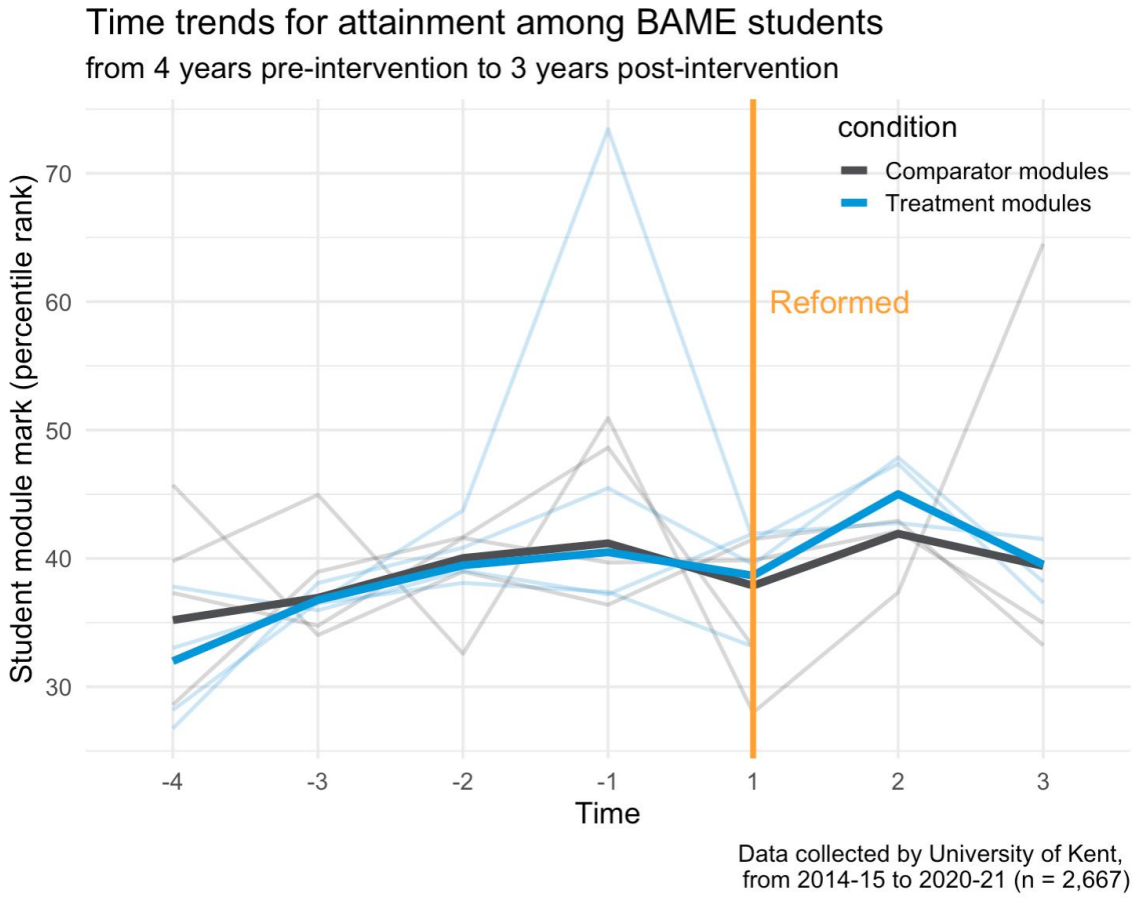
4.3. Results from regression analysis

Primary analysis

There is no evidence that the parallel trends assumption was violated in either of the four 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, after the Diversity Mark Initiative was piloted, we did not observe a significant impact of the Diversity Mark Initiative on improving attainment in terms of module mark percentile rank among BAME students. The average difference in attainment (2.0 percentile rank, 95% CI [-2.20, 6.21]) between reformed and unreformed modules post-intervention versus pre-intervention was not statistically significant, $p = 0.35$ (see Appendix 1 for full regression outputs). Figure 5 presents the trend of attainment year by year.

Figure 5. Time trends of student attainment among BAME students



Secondary analysis

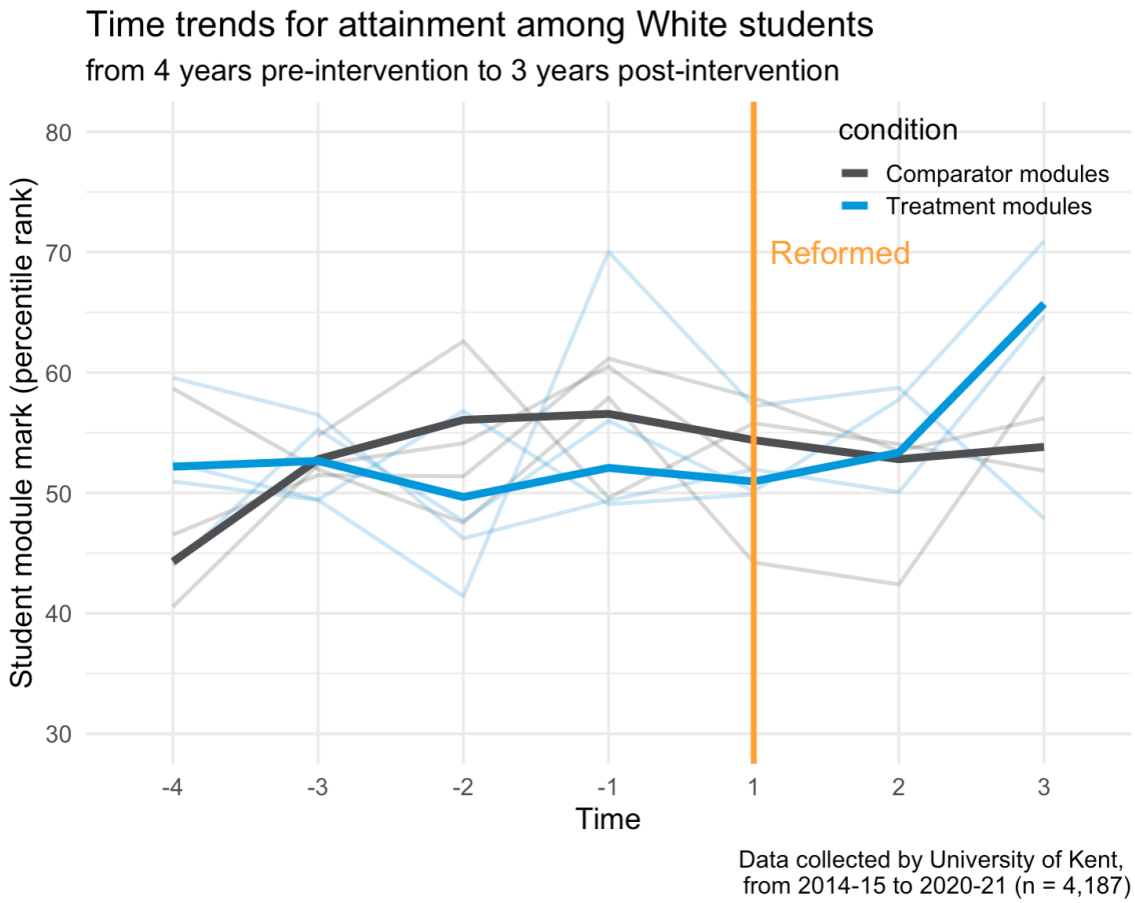
We did not check the parallel trends assumption formally for attainment among White students. But by visual examination (see **Figure 6**), the trends did appear to be adequately parallel up to 3 years prior to intervention, therefore we interpret the results as causal, though with less confidence as we did for the primary analysis.

Overall, we observed a marginally positive effect of the Diversity Mark Initiative on improving attainment among the White students. The average difference in attainment between reformed and unreformed modules post-intervention versus pre-intervention was 3.45 percentile rank, (95% CI [-0.13, 7.03], $p = 0.06$) (see **Appendix 1** for full regression outputs). Even though we were less confident in concluding the effect of the Diversity Mark initiative among White students, we consider it unlikely that it had lowered attainment among White students.

When we examined the effect of intervention year by year, we saw that the effect was strongest in the third year post-intervention, which corresponds to the academic year 2020/21, when many modules were delivered online rather than in person. We note that some studies have observed the racial awarding gap widened as a result of the pandemic[9]. However, we have no reason to believe that online learning affected the reformed and comparator modules differently - hence we remain agnostic about the reason why White students saw a sharper improvement in their percentile mark in year 20/21 in treated modules compared to comparison modules.

[9] Lally, C., & Bermingham, R. (2020). *COVID-19 and the disadvantage gap*. Retrieved from https://post.parliament.uk/covid-19-and-the-disadvantage-gap/

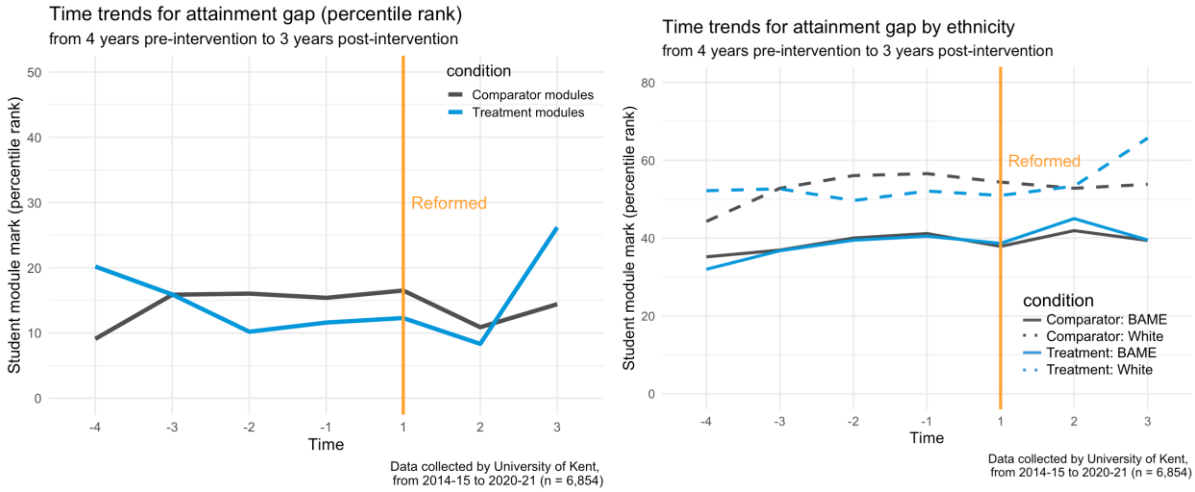
Figure 6. Time trends of student attainment among White students



Exploratory analysis

To understand the awarding gap between BAME and White students, we visualised the time trends of the attainment in terms of percentile rank (Figure 7) and % of students awarded an upper second class and above (Figure 8) for both ethnic groups.

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



As shown in **Figure 7, left panel**, in year 4 pre-intervention, the awarding gap between White and BAME students was narrower among the comparator modules (dark grey line) than that among the reformed modules (blue line), although the trend reversed since then and appeared to be parallel up to 2 years pre-intervention. Post intervention, the awarding gap seemed to have shrunk among all modules in the second year post-intervention (so in a similar way across reformed and unreformed modules), but then it widened again in the third year, particularly among the reformed modules

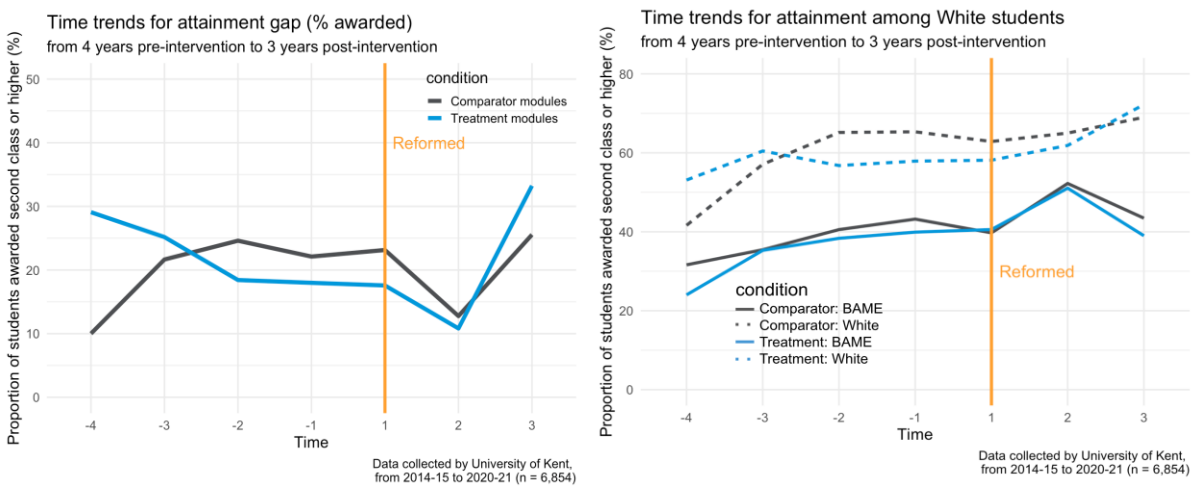
Figure 7, right panel can shed some light on why the gap increased more among reformed modules than among non-reformed modules in the third year after the intervention took place. This figure shows attainment trends among BAME students (solid lines) and White students (dashed lines) in reformed and unreformed modules (blue and dark grey lines respectively). White students had higher attainment than BAME students throughout the evaluation period. Among BAME students (solid lines), the average percentile rank of module mark hovered around 40 three years pre-intervention and remained fairly stable post-intervention. Among White students (dashed lines), the average percentile rank fluctuated between 45 and 55 up to the second year post-intervention and then increased substantially in the third year. This increase in the third year post-intervention was also more pronounced among the reformed modules (blue dash line) than the unreformed modules (dark grey dash line).

To further understand the awarding gap between BAME and White students, we also visualised the time trends of the attainment and awarding gap in terms of the proportion of students awarded upper second class and above (see **Figure 8**).

Overall, the patterns are fairly consistent. As shown in **Figure 8, left panel**, in year 4 pre-intervention, the awarding gap was narrower among the comparator modules (dark grey line) than that among the reformed modules (blue line), although the trend reversed since then and appeared to be parallel up to 2 years pre-intervention. Post intervention, the awarding gap seemed to have shrunk among all modules in the second year post-intervention, but then it widened among the reformed modules in the third year. And the widening was more pronounced among the reformed modules than among the comparator modules (36% vs. 26%).

Similarly, **Figure 8, right panel** illustrates what might have driven the increased award gap among the reformed modules (blue lines) than among non-reformed modules (dark grey lines) in the third year post-intervention. The widening award gap in the third year might be explained partially by White students' higher attainment among the reformed modules (blue dash line) and BAME students' lower attainment among the reformed modules (blue solid line).

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



While the widening gap in the third year post-intervention (as shown in **Figure 7 and 8, right panel**) seems to suggest that the intervention might have backfired, it is worth noting that the gap (especially the percentile rank gap) was chiefly driven by the higher attainment among White students rather than by lower attainment among the BAME students. Besides, we advise caution in interpreting this effect as causal or permanent, as it was observable only for a single year (which was heavily impacted by the COVID-19 pandemic). Moreover, the reformed modules also saw a higher proportion of enrolled BAME students post-intervention (see Table **5.3**), which might have contributed to the relative higher attainment among the White students.

5. Discussion

Overall, we didn't find conclusive evidence supporting the effectiveness of the Diversity Mark Initiative in reducing the racial awarding gap among the SSPSSR students at the University of Kent. However, we also didn't find evidence that suggests it might backfire and the observed trend in BAME students' attainment before and after the initiative, though not significant, was positive.[10]

This study has a few limitations that might affect the interpretation of the results. First, we did not have an objective quantification of to which extent the modules were reformed as we partially relied on course instructors' self-reported data when categorising which modules were reformed, therefore it was possible that some modules were less reformed than others. Second, there might be some spill-over effects as students might simultaneously have attended both reformed and unreformed modules, which could have diluted the treatment effects[11]. Third, the modules we included for analysis were not representative as we only included 4 reformed and 4 comparator modules from SSPSSR that had good historical data.

The above-mentioned limitations and other contextual factors might have constrained the generalisability of the findings. For example, as the sample modules were all drawn from social science courses that were not textbook-driven, we believe the findings would not generalise to science modules that rely more on textbooks[12]. Moreover, one module (Module 20) was reformed in the Summer of 2021, when the evaluation methods of modules (e.g. exams) might have been affected by the COVID-19 pandemic. We have no reason to believe this affected the treated modules any differently from the untreated modules (which is all we need to assume thanks to the difference-in-difference design) but it may impact the generalisability of the findings.

Nevertheless, we think this study is a useful addition to the evidence base as there was very limited empirical evidence in the effect of diversifying curriculum in reducing the racial awarding gap and it uses a quasi-experimental method that is the most feasible alternative to an Randomised Controlled Trial (RCT) in generating causal evidence[13]. Therefore, we consider the initiative an innovative approach to address the racial awarding gap that is worth further testing.

---

[10] See note 5 for details.

[11] One might also argue that the spill-over could have strengthened the treatment effect if a student prefers the treated modules after attending both treated and untreated modules, which makes the curriculum diversity difference more salient.

[12] This is unavoidable within the scope of this study, since the intervention *itself* does not have a clear generalisation to textbook-led courses.

[13] Hopkins, A., Breckon, J., & Lawrence, J. (2020, January 18). *The experimenter's inventory: a catalogue of experiments for decision-makers and professionals*. Retrieved from www.alliance4usefulevidence.org/join

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

	(1) BAME students	(2) White students
(Intercept)	37.68 **	52.01 **
	CI [32.61, 42.75], p < 0.001	CI [47.28, 56.73], p < 0.001
Post Intervention: Yes ×	**2.00**	**3.45 +**
Reformed: Yes	**CI [-2.20, 6.21], p = 0.35**	**CI [-0.13, 7.03], p = 0.06**
Post Intervention: Yes	-5.97 *	-9.01 **
	CI [-11.37, -0.57], p = 0.03	CI [-13.45, -4.57], p < 0.001
Reformed: Yes	-1.27	-1.23
	CI [-4.05, 1.50], p = 0.37	CI [-3.46, 1.00], p = 0.28
Year: 15-16	2.78	3.22
	CI [-2.14, 7.69], p = 0.27	CI [-1.10, 7.54], p = 0.14
Year: 16-17	4.49 +	1.35
	CI [-0.30, 9.27], p = 0.07	CI [-2.67, 5.37], p = 0.51
Year: 17-18	6.94 **	4.32 *
	CI [2.35, 11.52], p < 0.001	CI [0.38, 8.26], p = 0.03
Year: 18-19	11.22 **	10.64 **
	CI [5.19, 17.25], p = 0.00	CI [5.45, 15.83], p < 0.001
Year: 19-20	14.61 **	9.76 **
	CI [8.45, 20.77], p = 0.00	CI [4.68, 14.84], p < 0.001
Year: 20-21	10.18 **	14.96 **
	CI [2.99, 17.37], p = 0.01	CI [8.80, 21.11], p < 0.001
Module Pair 2	1.69	-1.45
	CI [-1.67, 5.05], p = 0.33	CI [-4.44, 1.54], p = 0.34
Module Pair 3	0.21	-0.25
	CI [-2.94, 3.36], p = 0.90	CI [-2.92, 2.41], p = 0.85
Module Pair 4	-4.25 *	-6.03 **

	CI [-8.29, -0.20], p = 0.04	CI [-9.25, -2.80], p < 0.001
Sex: Male	-6.15 **	-5.34 **
	CI [-8.83, -3.48], p < 0.001	CI [-7.42, -3.25], p < 0.001
Campus 2	-2.99	0.90
	CI [-6.74, 0.77], p = 0.12	CI [-2.74, 4.54], p = 0.63
N	2667	4187
R2	0.02	0.02

** p < 0.01; * p < 0.05; + p < 0.1.

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

	BAME students
(Intercept)	39.48 **
	CI [26.91, 52.05], p < 0.001
Year: 15-16	-2.83
	CI [-15.12, 9.46], p = 0.65
Year: 16-17	3.37
	CI [-5.32, 12.06], p = 0.45
Year: 17-18	4.15
	CI [-8.13, 16.42], p = 0.51
Year: 18-19	10.28
	CI [-3.68, 24.23], p = 0.15
Year: 19-20	13.62 +
	CI [-0.98, 28.22], p = 0.07
Year: 20-21	9.49
	CI [-6.19, 25.18], p = 0.24
Module Pair 2	1.71
	CI [-1.67, 5.08], p = 0.32
Module Pair 3	0.32
	CI [-2.86, 3.49], p = 0.85
Module Pair 4	-5.16 +
	CI [-11.18, 0.87], p = 0.09
Year of intervention: 4-year-pre-intervention	-0.72
	CI [-12.88, 11.45], p = 0.91
Year of intervention: 3-year-pre-intervention	3.07
	CI [-5.50, 11.64], p = 0.48
Year of intervention: 2-year-pre-intervention	-0.29
	CI [-8.55, 7.98], p = 0.95
Year of intervention: post-intervention	-7.01
	CI [-15.47, 1.44], p = 0.10

Reformed: Yes	-1.28
	CI [-6.31, 3.76], p = 0.62
Sex: Male	-6.14 **
	CI [-8.82, -3.46], p = 0.00
Campus: 2	-2.83
	CI [-6.60, 0.93], p = 0.14
4-year-pre-intervention × Reformed: Yes	**-2.01**
	CI [-10.05, 6.02], p = 0.62
3-year-pre-intervention × Reformed: Yes	**1.05**
	CI [-6.15, 8.25], p = 0.78
2-year-pre-intervention × Reformed: Yes	**0.29**
	CI [-6.85, 7.44], p = 0.94
Post-intervention: Reformed: Yes	2.05
	CI [-3.85, 7.94], p = 0.50
N	2667
R2	0.02

** p < 0.01; * p < 0.05; + p < 0.1.