REPORT ON AWARDING GAPS AND ASSESSMENTS

December, 2025

Gabriella Cagliesi
Business School, University of Sussex

Deyu Ming
School of Management, University College London

Susan Smith
School of Management, University College London

Valeria Terrones
Business School, University of Sussex

Luigi Ventimiglia
School of Economics and Finance, *Queen Mary University of London*

This report was developed through a cross-institutional collaboration led by the University of Sussex, with key contributions from Queen Mary University of London and University College London (UCL) **and funded by the QAA through its Collaborative Enhancement Projects scheme.**

Contents

Index of tables:

Index of Figures:

1. Executive Summary

This cross-institutional study examines how ethnic attainment patterns in UK higher education responded to changes in assessment design and learning environments across the Pre-COVID, COVID, and Post-COVID periods. Using large administrative datasets from the University of Sussex, Queen Mary University of London, and a case study from University College London, the analysis identifies both shared sector-wide phenomena and discipline-specific dynamics. The results demonstrate that awarding gaps are not fixed but vary systematically with assessment structures, institutional practices, and wider post-pandemic conditions.

Three central findings emerge across institutions. First, during the COVID period, ethnic attainment gaps narrowed or disappeared in several contexts, particularly in structured digital timed assessments. At Sussex, all non-Black gaps closed; at QMUL, gaps closed across all ethnic groups in Economics; and the UCL case study similarly showed improved convergence in online formats. This convergence suggests that emergency assessment conditions—greater scaffolding, simplified formats, and reduced high-stakes pressure—temporarily lowered structural barriers.

Second, when institutions transitioned back to stable assessment regimes, attainment gaps widened sharply. This divergence was most evident for Black students at all three institutions, reappearing across coursework and final assessments, and in both quantitative and non-quantitative modules. At Sussex, gaps re-emerged for all Black students across the wider Business School; at QMUL, the largest penalties appeared in extended written and untimed digital exams; and UCL reported similar challenges in open-ended, self-managed formats. These patterns indicate that the return to autonomous, high-stakes assessment environments disproportionately affected some groups.

Third, assessment type played a critical role. Extended, open-ended formats—essays, reports, take-home papers—generated the most pronounced post-COVID disparities, whereas structured timed assessments continued to show smaller or no gaps. Differences in quantitative performance shaped White–Asian gaps, but did not drive the widening White–Black gap. Across institutions, final assessments consistently posed greater challenges than coursework, particularly where weightings were high and time-management demands substantial.

Across all sites, post-COVID performance declined for all ethnic groups, suggesting a broader structural shift, including disrupted pre-university preparation, altered study habits, and post-pandemic pressures such as commuting, cost-of-living constraints, and reduced academic confidence. Intersectional factors reinforced vulnerability, with mental-health-related disabilities predicting sustained disadvantage at both Sussex and QMUL, and with evidence of post-COVID socioeconomic penalties for specific subgroups, such as FSM males at the University of Sussex.

Taken together, the results show that awarding gaps are highly sensitive to assessment design and post-pandemic study conditions. Equity improved most when assessments were structured, clearly guided, and time-limited, and worsened when open-ended, high-stakes, and self-regulated tasks were reintroduced. These patterns are consistent across three institutions with different student profiles and curriculum structures, underscoring the need for careful alignment between assessment formats, student support, and the structural conditions shaping learning after COVID-19.

2. Introduction

Attainment gaps by ethnicity persist in UK higher education and may be sensitive to how learning is assessed. The pandemic precipitated rapid changes to assessment conditions (for example, remote or open-book examinations), creating an opportunity to examine attainment patterns across different assessment modes (coursework, examinations, mixed of coursework and examinations) under differing operational contexts.

This report examines patterns of ethnic attainment in UK higher education and considers how assessment types and wider structural changes—particularly those linked to the COVID-19 pandemic—interacted with student background characteristics. Using administrative data from three universities: University of Sussex (UoS), Queen Mary University of London (QMUL), and University College London (UCL); the analysis explores how assessment types, timing, and institutional practices were associated with outcomes for different ethnic groups across three periods: Pre-COVID (2018–19), During COVID (2020–21), and Post-COVID (2022–23). The central aim is to understand whether awarding gaps changed across these phases and to identify assessment-related factors that may have contributed. This report is complemented with qualitative data with focus groups in these three universities (results are reported in a separate document).

3. Chapter 1: Methodology

3.1 Analytical Approach

This part of the report builds on earlier year-by-year analysis (Report 1). The earlier work highlighted broad trends across programmes and identified areas requiring further investigation. In this report, we use pooled mixed-effects models to provide a more integrated view of changes across the three periods and to support institution-level reflection on assessment practice, equity, and student support.

Our analysis followed a two-stage process. The first stage involved exploratory, year-by-year examination of ethnic attainment across a wide range of programmes: Economics, Economics and Finance, and Economic and Management at QMUL; and Economics, Accounting and Finance, and Management at UoS. This stage provided an overview of institutional patterns and helped identify emerging issues.

The second stage extended this work by pooling all three years into a single dataset. We first replicated the analysis for the same set of programmes reviewed in the exploratory stage. We then conducted a more focused analysis restricted to BSc Economics students (including placement variants).[1] Focusing on Economics enables clearer comparison between institutions because the discipline follows similar national subject benchmarks and assessment expectations. This focus also strengthens the interpretation of results within a quantitative social-science context, where mathematical preparation and assessment design are known to play a substantial role in shaping outcomes.

To provide further context, we also draw on the broader datasets to compare Economics with other Business School disciplines at UoS and to compare straight Economics with joint degrees at QMUL. These internal comparisons are not intended as cross-institutional matches but offer useful insight into whether patterns seen in Economics reflect wider trends within

[1] These two programmes share an identical on-campus curriculum and assessment structure; the only distinction is that some students complete a year-long industry placement between Levels 5 and 6. In the UK, an undergraduate degree is a level 6. Level 4, 5 and 6 refer to the first, second and third academic years, respectively.

each institution. For this reason, Economics remains the main focus of the narrative, with brief institutional comparisons presented at the end of each section.

This disciplinary focus strengthens comparability but naturally limits generalisability to other areas of the curriculum. Patterns of ethnic attainment may differ in less quantitatively intensive programmes, professional disciplines, or creative subjects. However, Economics provides a useful case study: it is offered widely across the sector, attracts diverse cohorts, and prepares students for competitive labour-market pathways where equity of opportunity is a key concern. Future work may extend this type of analysis to additional subject areas.

3.2 Rationale for the Pooled Modelling Strategy

Pooling data across the three periods provides several advantages for institutional evaluation. It enables formal testing of differences across time and of interactions between ethnicity and assessment period. It also accounts for student overlap between periods by modelling within-student correlation through random effects. Finally, pooling maximises the use of the available data, increasing the reliability of inferences. This approach supports more robust conclusions about changes in awarding patterns across assessment phases.

3.3 Sample Description (UoS and QMUL)

To ensure a like-for-like comparison between UoS and QMUL, the sample selection followed a discipline-based logic rather than simply using all available assessment records. Three departments within the Business School—Economics, Accounting & Finance, and Management—were selected because they are the closest disciplinary equivalents to the programmes included in the QMUL analysis.

Within these departments, we restricted the dataset to single-honours degrees. This avoids "cross-contamination" arising when joint-honours students take compulsory modules delivered by other Schools (e.g., Mathematics, Law, Psychology), which differ in assessment design, marking practices, and quantitative requirements. Focusing on single-honours pathways ensures that most assessments are produced within the same disciplinary context.

While this strategy greatly improves comparability across institutions, heterogeneity cannot be fully eliminated**.** Differences in curriculum design, accreditation requirements, cohort composition, and institutional assessment cultures inevitably remain. However, the sampling approach minimises these sources of variation and ensures that the results reflect genuine disciplinary and institutional patterns rather than confounding effects of mixed programme structures.

Tables 1 and 2 provide an overview of the student populations included in the analysis. Table 1 summarises characteristics for the broader set of Business School programmes (Economics, Accounting & Finance, and Management), enabling comparison across institutions at a wider level. Table 2 presents the same information for the Economics-only subsample, which is the primary basis for cross-institutional comparison in this report. More detailed descriptive statistics are provided in Appendix A (UoS) and Appendix B (QMUL).

Table 1: Sample Characteristics Comparison (Selected Programmes)

	QMUL	UoS	Notes
Programme	Selected Programmes (Econ, Econ and Finance, Econ , Finance and Management)	Selected programmes (Finance, Econ and Management) of the 3 Departments of the Business School	Partly matched
Entry requirements	AAA with A in math	ABB (pre-COVID) BBB (2020-21, 2022-23) Not A levels Maths[2]	Different
Time periods	2018-19, 2020-21, 2022-23	2018-19, 2020-21, 2022-23	Comparable
Assessment changes	Traditional → Online → Traditional	Traditional → Online → Online & traditional	Key difference
Ethnic composition	Asian 65%, White 16%, Black 10%, Mixed 4;% Other 5%(*)	Asian 14%, White 67%, Black 8%, Mixed 8%, Other 3%(**)	Ethnic profile differs: QMUL Asian-majority.
Sample size	17,941 observations (736 students of whom 211 overlapped across periods)	46,276 observations (2111 students of whom 775 overlapped across periods)	-

(*) QMUL: Give the small sample sizes for "Mixed and Other" ethnic minority students, statistical comparisons involving these groups should be interpreted with caution due to limited statistical power.
(**) UoS: Give the small sample size for "Other ethnic" minority students, statistical comparisons involving this group should be interpreted with caution due to limited statistical power.

Table 2: Sample Characteristics Comparison (Economics)

	QMUL	UoS	Notes
Programme	BSc Economics (+ placement variant)	BSc Economics and BA Economics (+ placement variant)	Matched
Entry requirements	AAA with A in math	ABB (pre-COVID) BBB (2020-21, 2022-23) Not A levels Maths[3]	Different
Time periods	2018-19, 2020-21, 2022-23	2018-19, 2020-21, 2022-23	Comparable
Assessment changes	Traditional → Online → Traditional	Traditional → Online → Online	Key difference
Ethnic composition	Asian 65%, White 15%, Black 11%, Mixed 4;% Other 5%(*)	Asian 12%, White 66%, Black 9%, Mixed 9%, Other 4%(**)	Ethnic profile differs: QMUL Asian-majority.
Sample size	12,689 observations (497 students of whom 141 overlapped across periods)	15,134 observations (655 students of whom 252 overlapped across periods)	Comparable

(*) QMUL: Give the small sample sizes for "Mixed and Other" ethnic minority students, statistical comparisons involving these groups should be interpreted with caution due to limited statistical power.
(**) UoS: Give the small sample size for "Other ethnic" minority students, statistical comparisons involving this group should be interpreted with caution due to limited statistical power.

[2] It returns to ABB since the academic year 2023-2024.
[3] It returns to ABB since the academic year 2023-2024.

3.4 Modelling Strategy

The pooled analysis follows the same overall approach used in the year-by-year models. We first examined all assessments taken by students within each academic year, controlling for whether the assessment took place during teaching weeks or in the final exam period. This provided an overview of attainment across the full range of assessment activity.

We then estimated two additional models to capture the distinct contribution of different assessment types:

- In-Term Assessment Model: Focuses on assessments completed during the teaching term and disaggregates results by assessment format (e.g., essays, problem sets, presentations, interactive tasks).
- Final-Exam Model: Examines only end-of-module assessment formats, which generally carry the highest weighting in overall module marks.

Across all three models, we applied a consistent set of control variables relating to student characteristics, module attributes, and assessment factors. These covariates are summarised in Table 3.

Table 3: Variables Included in Regression Models

Category	Variables
Student Demographics	Gender Ethnicity Socioeconomic background Disability status Placement year enrolment
Assessment Context	Time period (Pre-, During-, Post-COVID) Assessment timing (in term vs. final) Term of delivery (Term 1 or Term 2) (QMUL only)
Module Characteristics	Quantitative module (binary indicator) Module level (4, 5, or 6) Core vs. optional module (QMUL only) Assessment type (in term and final exam models)

3.5 Model specification

A substantial proportion of students contributed assessment data across more than one COVID period (28% at QMUL and 38% at UoS). For example, a student enrolled in 2018–19 may have results in both the Pre-COVID and During-COVID phases, while a 2020–21 entrant may contribute data to the During- and Post-COVID periods. In addition, some students repeated or trailed modules, which also extended their assessment histories across years.

To account for this within-student correlation, we used mixed-effects linear regression models with random effects at the student level. This approach reflects the nested structure of the data, in which multiple module marks (level-1) are clustered within individual students (level-2). By modelling both within- and between-student variation, the mixed-effects structure provides more reliable estimates and avoids treating repeated observations as independent.

All models were estimated in Stata using the *xtmixed* command, and the same equation structure was applied across three specifications: (1) a full model using all available grades, (2) an in-term assessment model, and (3) a final-exam model. The in-term and final-exam specifications included assessment-type variables and relevant interaction terms. Full

estimation tables for all models are provided in a separate compendium submitted with this report.

Model specification was consistent across institutions, with limited adjustments to reflect local differences—for example, the presence of core modules at QMUL. This ensures comparability while allowing for institution-specific structures.

Fixed-effects structure

All models include the following fixed-effect controls:

- Ethnicity, time period, and their interaction
- Module-level characteristics (quantitative intensity, core status, FHEQ level)
- Student-level characteristics (gender, disability, occupation group)

Standard errors were clustered at the student level.

Random-effects structure

Both the QMUL and UoS models include a random intercept for students, capturing individual differences in baseline attainment that may reflect unobserved factors such as prior preparation or motivation. Random slopes were introduced where appropriate to allow students to vary in their response to structural features of the curriculum.

In particular, both institutional models include a random slope for module level. This allows for individual variation in academic progression across FHEQ Levels 4, 5, and 6, reflecting the fact that students move through the degree at different paces and may complete assessments at more than one level within the same academic year (for example, due to trailing).

All models were implemented in Stata using the *xtmixed* command, and the same equation structure was applied across three specifications. Standard errors were clustered at the student level. The Overall Model (Model 1) takes the following form:

$$\boldsymbol{M_{iay} = \beta'_1 SC_{iy} + \beta'_2 MC_{ay} + \beta'_3 AC_{ay} + \beta'_4 IT_{iay} + u_i + \varepsilon_{iay}} \quad (\mathbf{1})$$

where M_{iay} is the mark obtained by student "i"on assessment "a" in academic year ""y. The vectors of explanatory variables are:

- SC_{iy}: student characteristics (e.g. ethnicity, gender, FSM eligibility, disability),
- MC_{ay}: module characteristics (e.g. progression level, quantitative modules, placement)
- AC_{ay}: assessment characteristics (e.g. assessment type, weight, in-term vs final),
- IT_{iay}: relevant interaction terms (e.g. ethnicity × assessment type, period × assessment mode)
- u_i the student-level random effect (with model-specific extensions where random slopes are included [4]
- ε_{iay} is the idiosyncratic stochastic error term (random effects and error term are assumed independent).

The random effects depend on each specification as it describes below. The remaining two models (model 2 and model 3) retain the same modelling structure but are estimated

[4] For clarity of presentation, the random-effects term $u_i$is written as a single component in equation (1). In practice, the model estimates a vector of student-specific random effects, including a random intercept and (where supported by the data) random slopes for variables such as academic level, quantitative intensity, and assessment period. These random effects jointly follow a multivariate normal distribution. Random effects were tested using likelihood ratio tests, information criteria, and examination of variance components, and were retained only when they demonstrably improved model fit.

separately for different assessment types, thereby controlling for in-term coursework and for final examinations.

Institution-specific random components

At QMUL, an additional random effect is included for whether a module is core. This reflects the expectation that students may perform differently in compulsory modules compared with optional ones. The same feature is not included for UoS due to the smaller number of core modules, particularly at Level 6.

Conversely, the UoS model includes a random component distinguishing assessments undertaken during term, final exams, and high-stakes (100%) final assessments. This separation captures meaningful variation in assessment conditions at Sussex but is omitted at QMUL because assessment formats are more homogeneous, particularly in the post-COVID return to traditional invigilated exams.

UoS models also include a random slope for quantitative intensity. Because A-level Mathematics is not a prerequisite, students vary in their preparedness for quantitative material. Allowing this slope to vary captures these differences in numeracy-related responsiveness and improves model fit. QMUL does not include this term due to A-level Maths entry requirement.

Finally, the time-period variable is treated as a random slope at UoS, reflecting the fact that a substantial share of students (38%) appears in more than one period and experienced meaningful changes in assessment conditions over time. This provides sufficient within-student variation to estimate period-specific random effects. At QMUL, only 28% of students overlap across periods and the post-COVID return to traditional examinations limits within-student variation between phases; therefore, time period is not included as a random slope.[5]

Together, these modelling choices ensure that both institutional models are grounded in a shared analytical framework while remaining sensitive to local assessment structures and curriculum design. These refinements strengthen the reliability of the estimates and supports meaningful comparison across assessment practices.

4. Chapter 2: Results for University of Sussex (UoS)

In the Sussex regression models, White students and the Pre-COVID period are used as reference categories. White students make up around 66% of the Sussex sample, making them the largest and most stable comparison group. Using the Pre-COVID period as the baseline allows changes during and after the pandemic to be interpreted relative to the last ‘normal’ assessment context. Presenting differences relative to White students follow standard practice in educational research and supports consistent interpretation of ethnic attainment patterns (see Tables 4–6).

[5] The random slope for progression level (L4–L6) is conceptually distinct from the random slope for time period, which captures variation in how individual students respond to the Pre-COVID, COVID, and Post-COVID phases. The two slopes therefore reflect different forms of within-student variation: normal academic development versus pandemic-related structural change.

Table 4: Ethnicity at University of Sussex (Economics)

BAME	Pre-COVID	COVID	Post-COVID	Total
White	69%	64%	65%	66%
Black	11%	9%	8%	9%
Asian	10%	15%	13%	12%
Mixed-heritage	8%	10%	10%	9%
Other ethnic group	2%	3%	4%	3%
Total	100% (228)	100% (220)	100% (207)	100% (655)
Average number of assessments per student	16	18	16	16
Coursework (In-Term) Final Exam (<100%) Final exam 100% weight	50% 40% 10%	57% 37% 6%	54% 39% 7%	53% 39% 8%
(*) Give the small sample sizes for "Mixed and Other" ethnic minority students, statistical comparisons involving these groups should be interpreted with caution due to limited statistical power.				

The full regression outputs for all three models—overall performance, in-term assessments, and final examinations—are provided in a separate results compendium submitted with this report (Tables C.1–C.6). This section presents a summary of the key findings, focusing on how ethnicity interacts with time period and assessment type. As noted in the modelling strategy, all estimates draw on the full set of control variables and random-effects structures described earlier.[6]

4.1 All Assessment Types: Overall Model (Economics)

Figure 1 presents the predicted attainment trajectories for each ethnic group across the Pre-COVID, COVID, and Post-COVID periods, showing how within-group performance changed over time. These predictions are derived using the full sample's covariate distribution to ensure comparability across periods. The figure highlights the key visual patterns, which are then formally tested in the regression results reported in Tables 5 and 6.[7]

[6] Random effects were tested using likelihood ratio tests, information criteria, and examination of variance components, and were retained only when they demonstrably improved model fit. These slopes capture individual heterogeneity arising from differences in mathematical preparation, progression/trailing patterns, and multi-period observation.

[7] Estimates for Other ethnic groups (each representing 4% of the sample) should be interpreted with caution due to small sample sizes and limited statistical power. We excluded this group from Tables 5-6 to focus on the four main groups (White, Black, Asian, and Mixed Ethnicity groups).

Figure 1: Ethnic Attainment Trajectories Across COVID Phases (Economics)

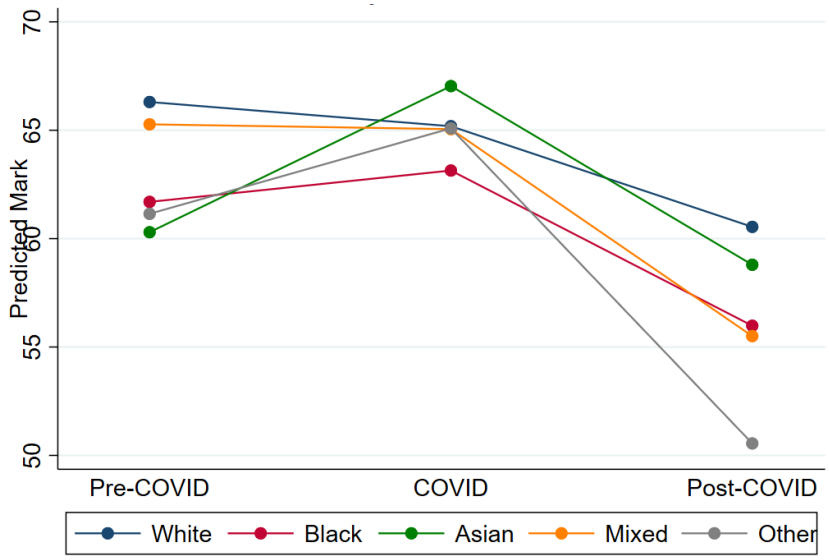


Figure 1 shows that attainment patterns shifted noticeably across the three COVID phases. During the COVID year, most ethnic groups either maintained stable performance or showed small improvements. Asian students, in particular, reached their highest predicted marks during this period, suggesting that digital online assessment conditions may have aligned better with their strength. The graph indicates:

COVID period (2020–21)

- White and Mixed-heritage students: performance remained broadly stable.
- Asian and Black students: modest improvement in predicted marks.
- Asian students: highest attainment recorded across all three phases

These patterns indicate that the immediate response to the pandemic did not widen ethnic attainment gaps. In fact, for some groups, outcomes improved modestly.

In contrast, the Post-COVID period shows a deterioration in attainment for all groups. These drops are not simply a return to pre-pandemic levels—they represent a clear decline in predicted performance across the board. The fact that every group was affected, though to different degrees, suggests broader structural influences linked to cohort characteristics or the post-pandemic academic environment.

Post-COVID period (2022–23)

- White students: decline of ~5 points.
- Asian students: return to pre-COVID baseline.
- Black and Mixed-heritage students: largest declines (approx. 7–9 points).

Common pattern: all groups experienced lower predicted marks relative to both Pre-COVID and COVID phases.

Possible contributing factors

These post-COVID declines may reflect several delayed effects of pandemic-era disruption, including:

- cancellation of A-level exams and reliance on teacher-assessed grades.

- atypical or uneven pre-university learning experiences.
- temporary changes to entry requirements.
- differences in readiness and transition into university-level study.

Notably, these impacts did not appear during the pandemic year itself but became more pronounced once standard teaching and assessment resumed. This delayed effect highlights the importance of sustained monitoring and support as students adjust to post-pandemic expectations.

Table 5: Within Group differences (Asian, White and Black students) (Economics)

BAME	Period		
	Covid vs. Pre-Covid	Post-Covid vs. Covid	Post-Covid vs. Pre-Covid
White	-0.431(ns)	**-4.688*** **	**-5.099*** **
Black	2.135(ns)	**-7.166*** **	**-5.051***
Asian	**7.428*** **	**-8.267*** **	-0.839(ns)
Mixed	0.467(ns)	**-9.574*** **	**-9.106*** *

Note: *** p<0.01, ** p<0.05, * p<0.10, (ns)= not statistically significant. Time columns show predicted change in marks for each ethnic group relative to their pe-COVID (2018-19) baseline.

Table 5 reports the predicted (and tested) within-group differences across periods, derived from the regression model.

The key patterns visible in Table 5 are:

1) White students
    - Little change during COVID
    - Clear and significant decline post-COVID
2) Black students
    - non-significant improvement during COVID
    - Sharp decline post-COVID (–7.2 vs. COVID; –5.1 vs. Pre-COVID).
3) Asian students
    - Large improvement during COVID (+7.4 points).
    - Largest post-COVID drop (–8.3 vs. COVID).
    - End up roughly back at the Pre-COVID baseline.
4) Mixed-heritage students
    - No significant change during COVID.
    - Severe post-COVID decline (–9.6 vs. Pre-COVID).

In other words, all groups decline post-COVID, but Black and Mixed-heritage students decline the most, widening the awarding gap. Thus, these within-group changes help explain why cross-group attainment gaps reopened after the pandemic, as reported in Table 6. We can observe the following:

- For Asian students, the COVID-period gains were temporary: the sharp post-COVID drop brings them back to their baseline.
- The White–Black gap re-emerge and the White–Mixed emerged not because White students improve, but because Black and Mixed-heritage students fall further.
- This supports the view that the transition out of the pandemic, rather than the pandemic year itself, coincided with the widening of ethnic differences.

Table 6: Ethnic Awarding Gaps Across COVID Phases (Economics)

	Pre-Covid Gap	Covid Gap	Post Covid Gap	COVID Achievement	Post-COVID Status
White vs Black	**4.609****	2.040 (ns)	**4.561***	**Gap Reduced**	**Gap Persists**
White vs Asian	**6.007*****	-1.851(ns)	1.747(ns)	**Gap Eliminated**	**Gap Narrowed**
White vs Mixed	1.031	0.133 (ns)	**5.039***	No Baseline Gap	**New Gap Emerged**

Note: *** p<0.01, ** p<0.05, * p<0.10. Gaps calculated as White mean minus other ethnicities; positive values indicate White advantage.
GAP: time-period shows absolute gaps at each time point (positive = White advantage)

The findings suggest that COVID represented a unique moment of equity when traditional gaps narrowed or disappeared, but this was temporary for most groups. The critical question is why this occurred: Was the adoption of new assessments (digital online timed and untimed assessments), support structures (financial aid, mental health services, academic flexibility), or universal disruption that narrowed the awarding gaps? The reversion to baseline for White-Black inequality and the emergence of new Mixed-Heritage disadvantage suggest that without active intervention to address disparities, institutional systems may revert to reproduce pre-existing patterns of inequality.

Beyond Ethnicity: Gendered Socioeconomic Inequality. Reversal of Advantages.

While the ethnicity analysis highlights a widening of attainment gaps in the post-COVID period, it does not capture the full complexity of disadvantage. To explore how multiple dimensions of identity intersect, we examine the combined effects of gender and socioeconomic status (measured by Free School Meal (FSM) eligibility). This intersectional perspective reveals a significant post-pandemic shift: prior to COVID, disadvantaged female students faced greater challenges, but after the pandemic it was FSM males who emerged as the most adversely affected group.

Table 7 examines how gender and socioeconomic status intersect across the three periods. To make these patterns clear, we separate the analysis into within-group changes over time and across-group gender differences.

Table 7: Gender Gaps and FSM Effects Across Time (Economics)

Time Period	Gender Gap Non-FSM	Gender Gap FSM	FSM Effects on Males	FSM Effects on Female
Pre-COVID	**4.22***** (p=0.001)	-5.970(ns) (p=193)	+2.730 (ns) (p=0.273)	**-7.280*** (p=0.061)
COVID	1.780(ns) (p=0.299)	-1.330(ns) (p=719)	0.220(ns) (p=0.893)	-2.890(ns) (p=432)
Post-COVID	**4.00**** (p=0.034)	8.580(ns) (p=0.146)	**-7.420*** (p=0.095)	-2.84(ns) (0.509)

Note ***p<0.01, ** p< 0.05; * p<0.10, (ns)= not statistically significative
FSM Positive = FSM students outperform non-FSM, using dydx (marginal effects)
FSM negative = FSM disadvantage; using dydx (marginal effect)
Gender gap: Positive = Female advantage; negative = Male advantage (using predicted performance via margins).

Within groups, FSM males and FSM females show contrasting trajectories. Before the pandemic, FSM penalties were more evident for females, while FSM males performed similarly to their non-FSM peers. During COVID, these differences narrowed for both genders. In the post-COVID period, however, FSM males experienced a significant decline relative to non-FSM males, becoming the most disadvantaged group. FSM females did not exhibit a comparable post-COVID penalty.

Across groups, gender differences were concentrated among non-FSM students. Non-FSM females significantly outperformed males' pre-COVID, the gap narrowed during COVID, and re-emerged afterwards. Among FSM students the gender gap was less stable due to small sample sizes, but the general pattern shows convergence during COVID and after. The absence of a statistically significant gender gap Post-COVID among FSM students does not imply that FSM disadvantage affects males and females equally.

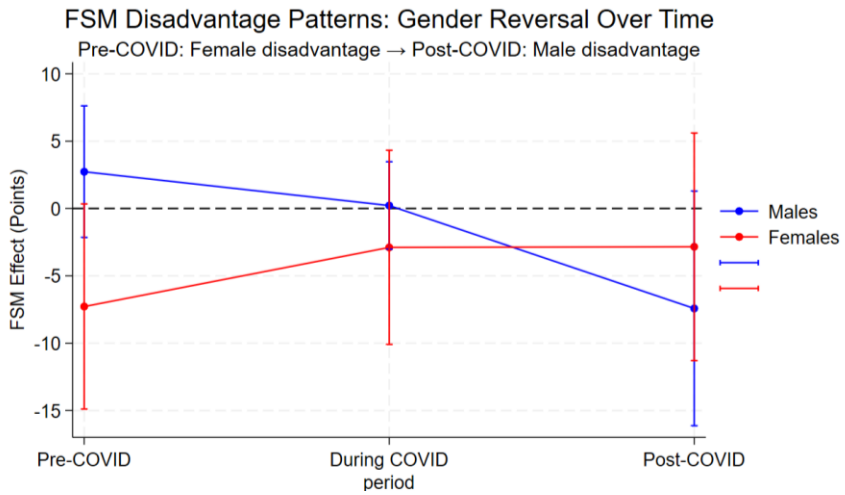
The key contrast comes from the FSM penalties within each gender. In the post-COVID period, FSM males show a large and statistically meaningful penalty relative to non-FSM males, whereas FSM females do not show a comparable penalty relative to non-FSM females. This indicates that socioeconomic disadvantage became particularly concentrated among male students, even though the gender gap within the FSM subgroup itself is not statistically significant possibly due to small sample sizes.

Taken together, these results show a reversal in the gendered pattern of socioeconomic disadvantage: whereas FSM females were more disadvantaged before the pandemic, the post-COVID recovery period saw FSM males emerge as the group facing the greatest challenges.

Figure 2 illustrates these patterns visually. The vertical axis shows the difference in predicted attainment between FSM and non-FSM students (negative values indicate lower performance for FSM students). Three features stand out:

- A small or moderate FSM penalty for females in the pre-COVID period.
- Convergence during COVID for both genders.
- A pronounced widening in the post-COVID period for males, with FSM males showing the largest socioeconomic-related decline across the entire study period

Figure 2: FSM Disadvantage Patterns: Gender Reversal Over Time (Economics)



A summary of all model estimates—including coefficients and p-values for assessment characteristics, progression level, disability categories, mental-health indicators, and module attributes—is provided in Table 13 at the end of this section. Although the discussion here focuses on Model 1, several of the patterns observed at this stage also recur in the in-term and final-exam models as follows:

Assessment structure: Performance is significantly lower in assessments held outside the teaching term (Final exams) and in high-stakes (100%) final exams, relative to in-term coursework. This indicates that assessment structure and stakes have a substantial influence on outcomes, likely reflecting differences in scaffolding, preparation demands, time pressure, or exam-related stress.

Disability indicators: with the exception of the mental-health category—which is associated with a notable performance penalty—other disability categories do not show statistically significant differences in attainment.

Module-level effects: follow a clear progression pattern. Students perform significantly less well in Level 5 modules relative to Level 4, suggesting that the transition into intermediate study presents specific challenges. Level 6 modules do not differ significantly from Level 4, implying that the main difficulty occurs at the mid-stage of the degree. Strongly quantitative modules are also associated with lower predicted marks, likely reflecting variation in students' numeracy backgrounds. Students enrolled on placement modules tend to achieve higher marks, consistent with the benefits of experiential learning or greater engagement among placement-track students. These structural factors provide an important backdrop for interpreting the ethnic-attainment patterns discussed earlier.

We also conducted quantile regression analysis to assess whether patterns of advantage and disadvantage vary across the performance distribution. Full estimates are available on request; the key findings are summarised here.

For Black and Mixed-heritage students, disparities were concentrated at the lower and upper quartiles. These gaps disappeared during the COVID period but re-emerged post-COVID, often more sharply at the performance extremes. Quantitative modules show a similar asymmetric pattern: large penalties for lower-achieving students, modest disadvantage at the median, and small gains among higher performers. FSM males showed no systematic penalty before or during COVID, but in the post-COVID period experienced significant disadvantages across all quartiles, with the largest effects at the lower end.

Taken together, these findings show that patterns of disadvantage are not static. They shift with broader contextual and institutional conditions. This challenges deficit-based interpretations of inequality and highlights the need for targeted support that not only assists students at risk of underachievement but also enables high-achieving students from underrepresented groups to maintain and extend their performance.

To explore the institutional dynamics behind these patterns in more depth, the next section examines the role of assessment type. We separate the sample into coursework and final-exam subsamples to identify the distinct effects of each on student outcomes.

4.2 In-Term Assessments: Coursework Model (Economics)

In this model, we restrict the analysis to coursework only (i.e., different types of coursework occurred during the teaching term), and we consider different types of assessments. Within this subset, we created five categories, using the university classification and terminology, to distinguish and capture differences in task structure: (i) short-timed assessment; (ii) essays; (iii) written report; (iv) technical task; and (iv) interactive and engagement-based task. Table 8 provides a description of the categories.

Table 8: Coursework Assessment Types and Descriptions (Economics)

Assessment Category	Description
Short Timed Assessments	Assessments completed within a limited time (1 or 2 hours. This category includes in class or out of class written tests (TST) (paper-based or online; computer-based examination (CEX) tests; multiple choice question (MCQ) tests.
Essays	Extended written responses, usually requiring structured argumentation based on existing literature. This category includes standard essays (ESS).
Written Reports	Structured written work based on analysis, research or reflection. This category includes reports (REP), project work (PRJ), and portfolios (POF).
Technical Tasks	Assignments focused on solving defined problems or producing technical outputs. This category includes problem sets (PRB); media (MED), software exercises (SOE).
Interactive and Engagement-Based Tasks	Activities involving verbal or group-based interaction and communication skills. This category includes presentations (GPN), oral exams (ORL), and collaborative group work (GWS).

This classification allows us to test whether the post-COVID divergence in attainment is specific to certain coursework formats, or whether the same dynamic is visible across all of them. The distribution of these assessments is presented in Table 9.

Table 9: Distributions of Assessments Across COVID Phases (Economics)

Assessment Category	Pre-COVID	COVID	Post-COVID	Total
Short Timed Assessments	576 (25%)	1,042 (31%)	989 (41%)	2,607 (32%)
Essays	685 (29%)	688 (20%)	660 (27%)	2,033 (25%)
Written Reports	139 (6%)	329 (10%)	300 (12%)	768 (9%)
Technical Tasks	346 (15%)	349 (10%)	156 (7%)	851 (10%)
Interactive and Engagement-Based Tasks	575 (25%)	979 (29%)	308 (13%)	1,862 (24%)
Total	2,321 (100%)	3,387 (100%)	2,413 (100%)	8,121 (100%)
Average number of assessments per student	8	10	9	9

Table 9 shows shifts in in-term assessment practices across the COVID phases. These changes were notable during the pandemic, although only some elements persisted afterwards. From the table, several patterns emerge:

1) The number of in-term assessments increased during the COVID period (from about 8 to 10 per student), reflecting the shift toward more continuous assessment during remote learning. However, this increase did not persist post-COVID, the volume returned to pre-pandemic levels.

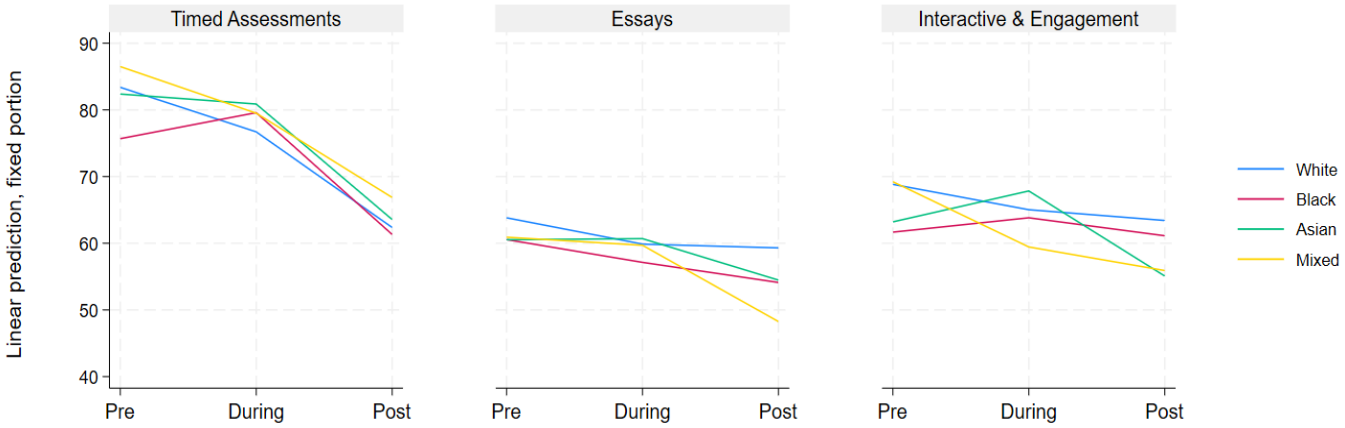
2) Changes in the *mix* of assessment types were more pronounced: Short-timed assessments and written reports rose steadily across all three periods, continuing to increase even after COVID; Technical tasks declined consistently over time; Essay-based assessments fell sharply during COVID and only partially recovered afterwards; Interactive and engagement-focused tasks rose substantially during COVID (28.9%) to support online teaching but fell sharply post-COVID (12.8%). Thus, some changes were pandemic-specific, while others indicate longer-term shifts in departmental practice.

Despite these fluctuations, the broad structure of in-term assessment was stable. Across all periods, roughly 80% of assessments fell into three categories: Short-timed assessments, Essays and Interactive/engagement tasks. This indicates that the overall distribution of assessment types remained broadly consistent, even if the balance within these categories shifted.

Because specific assessment types were not evenly distributed across periods, changes in student performance over time reflect a combination of the timing of assessments, and the types of assessments assigned in that period. Assessment and time effects therefore interact, rather than representing separate influences.

The next section examines predicted margins for the three dominant assessment types—Short-Timed Assessments, Essays, and Interactive/Engagement-Based Tasks. These categories together not only account for roughly 80% of in-term assessments but are also sufficiently represented across ethnic groups and periods. Their predicted values are shown in Figure 3.

Figure 3: Ethnic Attainment by Assessment Type (Economics)



By examining Figure 3, three main patterns emerge: First, in comparing assessment types, it is clear that essays and interactive engagement tasks show more moderate levels of performance compared to timed assessments, which record the highest average marks up to the COVID period. This trend is consistent with established marking practices, as assessments with clearly defined "correct" answers—such as multiple-choice or structured tests—tend to yield a broader distribution of higher grades. In contrast, essays and interactive

tasks, which typically involve more open-ended or subjective evaluation, produce more stable but lower average marks.[8]

Second, when examining changes over time, all assessment types display a decline in average marks, converging toward the mid-60s in the post-COVID period. This downward trend may reflect the broader academic profile of students entering university during and after the pandemic, including the use of predicted grades and lower entry requirements. The most pronounced decline occurs in short-timed assessments. One likely contributor is the increased use of these tasks after COVID, which may have intensified assessment load during teaching weeks and reduced students' preparation time. Essays—although lower scoring overall—typically involve longer preparation periods and may have been less affected by this compression of assessment schedules.

Third, comparing across student groups reveals that ethnic gaps are smallest in timed assessments. During and after COVID, digital timed coursework shows almost no awarding differences between ethnic groups. However, disparities persist in more interpretive assessment formats. Essay-based assessments continue to show lower performance, particularly for Mixed-Heritage students, and smaller gaps remain in interactive tasks. This suggests that while pandemic-related changes in assessment design may have narrowed gaps in structured formats, inequalities remain in tasks requiring extended writing, interpretation, or applied reasoning.

These patterns align with the attainment differences reported in Table 10. In the post-COVID period, no significant gaps appear for Asian students across any coursework type. Moderate gaps remain for Black and Mixed-Heritage students in written reports; for Mixed-Heritage and Other ethnic groups in essays; and for students in the Other group in technical and interactive tasks. Importantly, no gaps appear in short-timed assessments, reinforcing the view that residual disparities are concentrated in more open-ended formats.

Table 10: Ethnic Awarding Gaps Post-COVID by type of Coursework (Economics)

Awarding gaps	Short-timed	Essay	Written Reports	Technical Tasks	Inter. Eng. Tasks
White vs Black	+1.051 (0.807)	+5.186 (0.182)	**+12.442 **** **(0.031)**	+0.243 (0.962)	+2.288 (0.642)
White vs Asian	-1.194 (0.697)	+4.829 (0.133)	+1.342 (0.686)	+3.101 (0.488)	+8.328 (0.125)
White vs Mixed-Her.	-4.498 (0.295)	**+11.052 **** **(0.016)**	**+13.568*** **(0.085)**	+9.644 (196)	+7.503 (0.162)
White Vs Other	-2.373 (0.776)	**+11.029*** **(0.069)**	+12.448 (0.101)	**+10.679*** **(0.074)**	**+20.400**** **(0.022)**

Note: *** p<0.01, ** p<0.05, * p<0.10. Columns show ethnic attainment gap in the post-COVID period for each type of coursework assessment. P-value in parenthesis.

Finally, the coefficient estimates for the control variables in the coursework-only model remain highly consistent with those in the overall model. Gender, disability, and socioeconomic effects appear with similar magnitudes and directions, indicating that the underlying control-factor patterns are stable across specifications. Two changes are worth noting: Level 6 modules now show a four-point advantage over Level 4, and the quantitative-module penalty observed in

[8] Elsamanoudy et al., (2024). Islam et al., (2017). Scouller, K. (1998).

the overall model is no longer present[9]. Because the control-variable estimates are substantively consistent across models, we do not reproduce the full table here; the detailed results appear in Table 13 at the end of the section.

4.3 Out-of-term Assessments: Final Exams Model (Economics)

Final assessments underwent major transformation during the pandemic. Before COVID-19, almost all final exams (around 75%) were delivered in person as traditional Unseen Exams (UEX) under invigilated conditions. During the COVID period, this model was entirely replaced by digital online assessments—either timed (such as MCQs and structured tests) or untimed formats (such as essays or 24-hour take-away papers), depending on module design. Timed exams were intended to replicate the structure of in-person exams, whereas untimed assessments created a more flexible, open-book format.

In the post-COVID period, the Economics Department did not return to the pre-pandemic UEX system. Instead, digital assessments were kept, combining timed and untimed formats according to pedagogical needs and logistical constraints. This shift reflects a broader institutional move around flexibility, academic integrity, and inclusive practice.

Table 11 below summarises the distribution of final-assessment formats across the three periods.

Table 11: Final Exams Across Time (Economics)

Final Exams	Pre-COVID	Covid	Post-COVID	Legend
Timed Unseen Exam	79%	0	0	Traditional exams (UEX)
Timed Non -UEX Exam	5%	34%	50%	Including: Open Text Exams (OEX), Seen Exam (EXS), Test (TST) and Computer Based Exam (CEX) and Multiple-Choice Question (MCQ)
Untimed Exam	16%	66%	50%	Including: Essay (ESS), Project (PRG) and Take Away Paper (TAP)
Total	100% 2,312	100% 2.594	100% 2,050	Total final exams 6,956
Average number of final exams per student	8	8	7	7

Figure 4 illustrates predicted marks for final examinations across assessment formats and COVID phases. As expected, final exams show consistently lower attainment than coursework, reflecting their higher-stakes nature and stricter marking standards.

In Pre-COVID period Black students underperformed in traditional unseen exams (UEX) , recording the lowest mark in this format; Asian students performed similarly to their peers in UEX but scored lowest in untimed final assessments, such as take-home essays or extended submissions.

[9] The stronger time-period effects in the coursework-only model reflect the change in reference category: from all coursework to a specific category of coursework, in term timed exams—a format that shifted substantially across the COVID phases—the model captures larger period-to-period contrasts than when coursework (dummy out) is the baseline.

During the 2020–21 emergency shift to digital assessments, performance rose across all groups, and ethnic differences narrowed significantly—especially in digital timed exams, where no significant gaps were observed.

However, this convergence was short-lived. In the post-COVID period, ethnic disparities re-emerged, particularly for Black students, who scored lowest in digital untimed exams—now accounting for over 50% of all final assessments.

Overall, the figure shows a shift in the pattern of inequality:

- COVID-era digital assessment temporarily reduced gaps, particularly in structured timed formats.
- Post-COVID assessment practices—dominated by untimed digital exams—are associated with renewed ethnic disparities**,** especially affecting Black students.

Figure 4: Ethnic Attainment in Final Exams (Economics)

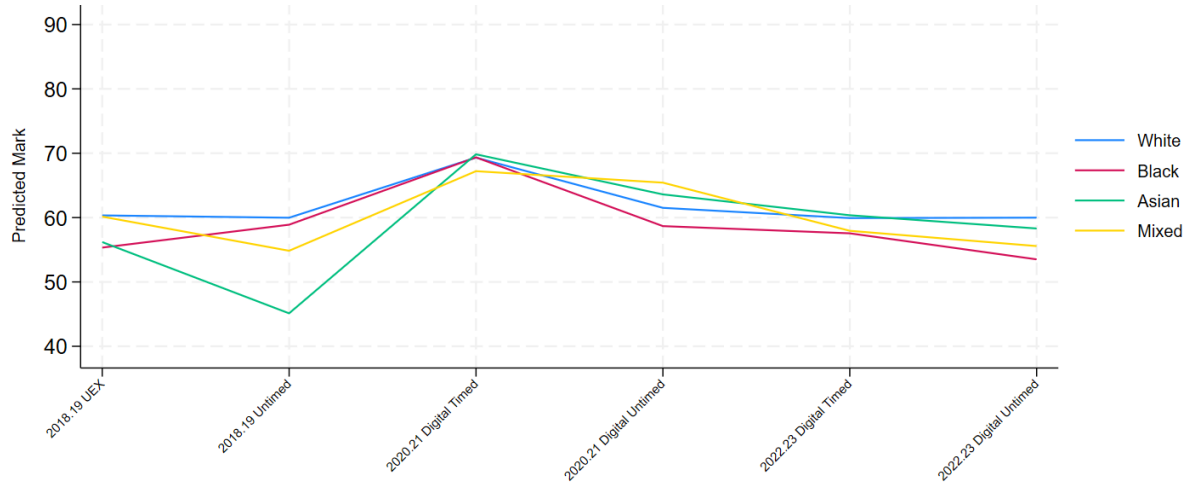


Table 12 quantifies these differences and the magnitude of the awarding gaps. While digital timed exams in 2020–21 show no statistically significant differences between ethnic groups, post-COVID digital untimed formats are associated with renewed and substantial disadvantage for Black students.

Untimed digital final assessments appear flexible but introduce several challenges that disproportionately affect some groups. Students must manage their own time, structure their responses, and complete high-stakes tasks in home environments that vary widely in terms of stability, quiet space, and competing responsibilities.

These findings suggest that assessment format plays a meaningful role in shaping ethnic disparities. The temporary equalisation observed during the COVID phase illustrates that alternative assessment designs can reduce structural inequalities—but also highlights the need to ensure that such improvements represent genuine equity gains rather than artefacts of assessment conditions. In the post-COVID context, deliberately designing final assessments with attention to differential impacts across groups may be essential for supporting more equitable outcomes.

Table 12: Ethnic Attainment Effects by Type of Final Exams Assessments (Economics)

Ethnic Group	2018-19 UEX	2018-19 Timed exams	2018-19 Untimed exams	2020-21 Timed exams (digital)	2020-21 Untimed exams(digital)	2022-23 Timed exams (digital)	2022-23 Untimed exams(digital)
White vs. Black	**5.000****	3.959	1.084	-0.035	2.836	2.369	**6.479****
	(p=0.027)	(p=0.440)	(p=0.618)	(p=0.990)	(p=0.316)	(p=0.500)	(p=0.034)
White vs. Asian	4.147	5.437	**14.858*****	-0.508	-2.094	-0.443	1.687
	(p=0.138)	(p=0.363)	(p=0.004)	(p=0.817)	(p=0.295)	(p=0.815)	(p=0.481)
White vs. Mixed	0.192	**-9.220*****	5.136	2.113	**-3.916****	1.947	4.411
	(p=0.922)	(p=0.007)	(p=0.223)	(p=0.351)	(p=0.029)	(p=0.304)	(p=0.115)
White vs. Other	7.874	3.307	17.702	2.786	0.031	**13.621****	**10.273*****
	(p=0.109)	(p=0.839)	(p=0.167)	(p=0.449)	(p=0.992)	(p=0.019)	(p=0.002)
Result: Gaps group	Black Disadvan.	Mixed Advant.	Asian Disadvan.	**No gaps**	Mixed Advant.	Other Disadvan.	Black and Other Disadv.

Note: *** p<0.01, ** p<0.05, * p<0.10. P-values in parenthesis.

The effects of controlling factors are consistent with the overall models, with stronger negative impact of quantitative model and a stronger impact of Level 6 progression. Disability, and socioeconomic have a negative impact on exam performance, for both female and male. Table 13 summarises the results of all models.

Table 13: Additional Model Effects - Control Variables and Interactions, all Models (Economics)

Variables	Overall Sample		Coursework		Final Exam	
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
A. ASSESSMENT & MODULE LEVEL EFFECTS						
Out of term (Final exams) vs. Coursework	**-3.697*****	0.000				
Out of Term (High stake FE) vs. Coursework	**-4.577*****	0.000				
Progression Level 5 vs. Level 4	**-8.341*****	0.000	**-4.699*****	0.000	**-3.303*****	0.001
Progression Level 6 vs. Level 4	-0.136	0.863	**4.087*****	0.000	**5.183*****	0.000
Quantitative Module vs. non-Quant.	**-1.640*****	0.000	0.023	0.968	**-1.974*****	0.000
Placement vs non-Placement	**4.698*****	0.000	**3.913****	0.017	**5.046*****	0.000
Time: COVID (2020)	-0.431	0.672	**-5.869*****	0.003		
Time: Post-COVID (2022)	**-5.099*****	0.000	**-20.255*****	0.000		
B. GENDER EFFECTS & SOCIOECONOMIC EFFECTS						
FSM (1 = FSM)	2.732	0.273	3.041	0.222	1.798	0.493
Female (1 = Female) vs Male	**4.222*****	0.001	**4.307*****	0.001	**4.378*****	0.005
FSM × Female	**-10.008****	0.032	**-10.128****	0.066	**-9.855****	0.028
FSM × Time: COVID (Male)	-2.508	0.311	-3.596	0.169	-2.442	0.349
FSM × Time: Post-COVID (Male)	**-10.153****	0.046	**-11.387****	0.043	**-8.335***	0.068

Female × Time: COVID	-2.438	0.230	-2.651	0.186	-2.552	0.296
Female × Time: Post-COVID	-0.218	0.920	-0.550	0.824	-0.526	0.823
FSM × Female × Time: COVID	6.896	0.210	7.742	0.220	6.831	0.198
FSM × Female × Time: Post-COVID	**14.585****	0.050	**18.032****	0.031	11.395	0.110
C. ETHNICITY EFFECTS						
Black vs. White	**-4.609****	0.014	**-7.703***	0.056	**-5.000****	0.027
Asian vs. White	**-6.007****	0.002	-1.010	0.796	-4.147	0.138
Mixed vs. White	-1.031	0.561	3.104	0.512	-0.192	0.922
Other vs. White	-5.159	0.189	**-6.819****	0.043	-7.874	0.109
Black × Time: COVID	2.565	0.331	**10.576****	0.033		
Black × Time: Post-COVID	0.047	0.988	6.651	0.254		
Asian × Time: COVID	**7.859*****	0.001	**5.180**	0.274		
Asian × Time: Post-COVID	4.26	0.133	2.204	0.654		
Mixed × Time: COVID	0.898	0.721	-0.306	0.959		
Mixed × Time: Post-COVID	-4.007	0.208	1.395	0.827		
Other × Time: COVID	5.062	0.318	**8.329***	0.091		
Other × Time: Post-COVID	-4.832	0.453	9.192	0.302		
D. DISABILITY EFFECTS						
Disability: Cognitive	-1.011	0.602	-0.047	0.983	-2.575	0.308
Disability: Mental Health	**-6.041*****	0.001	**-5.365*****	0.005	**-5.258*****	0.006
Disability: Multiple	0.735	0.739	0.876	0.685	1.335	0.632
Disability: Physical	-0.694	0.725	-0.154	0.940	1.513	0.364
Disability: Social	-1.689	0.799	-1.480	0.170	-2.631	0.561
E. RANDOM EFFECTS (SD)						
Student intercept	5.559		5.995		7.171	
Module slope (quant)	6.499		5.125		4.848	
Module level slopes (Level 4)	8.504		9.414		7.281	
Module level slopes (Level 5)	7.377		7.218		9.556	
Time period slopes (COVID)	6.388		5.204			
Time period slopes (Post-COVID)	8.531		10.331			
Out of term assessment slope	3.859					
Residual	17.726		17.906		15.011	
Note: *** p<0.01, ** p<0.05, * p<0.10.						

4.4 Conclusions of UoS (Economics)

The analysis shows that ethnic attainment gaps are not fixed but shift in response to assessment design, academic context, and wider structural conditions. Three broad conclusions emerge:

1) **COVID-19 temporarily reduced inequality.** During the pandemic emergency, attainment gaps narrowed or disappeared, particularly in digital timed final exams. This

suggests that certain assessment formats—especially short, structured, time-limited tasks—may reduce barriers for some groups.

2) **The convergence was not maintained post-COVID.** When assessment practices stabilised, substantial gaps re-emerged. Black students, in particular, experienced renewed disadvantages in longer, analytical assessment formats such as written reports and digital untimed final exams. Asian students, by contrast, no longer displayed the substantial pre-COVID disadvantages observed in earlier periods.

3) **Post-COVID performance declined across all groups.** Marks fell for every ethnic group after the pandemic. High-stakes, fully-weighted final exams were especially challenging, while coursework remained comparatively more stable.

Beyond ethnicity, intersecting sources of disadvantage were evident. Mental-health-related disabilities consistently predicted lower attainment. Quantitative modules were more challenging in final examinations but did not penalise students in coursework. Progression patterns showed a clear dip at Level 5 and recovery at Level 6. Gender interacted strongly with socioeconomic background, with FSM males experiencing the most pronounced post-COVID challenges.

Several limitations should be acknowledged:

- **Generalisability is limited.** Findings are based on one institution and one disciplinary cluster. Results may differ for programmes with different pedagogical structures or less quantitative content.
- **Small subgroup sizes.** Some ethnic groups (“Mixed”, “Other”) have small samples in certain assessment formats, reducing statistical power.
- **Focus on summative assessment.** The quantitative analysis examines summative assessment outcomes and does not capture day-to-day learning or formative processes. However, this is partially addressed by the project’s qualitative strand, which includes student focus groups on assessment preferences, perceived fairness, and experiences with online and post-COVID formats. These insights complement the quantitative findings.
- **Intersectionality requires deeper investigation.** FSM males and students with mental-health-related disabilities emerge as particularly vulnerable groups. Further research is needed to understand why they faced disproportionate post-COVID challenges.
- **Risks in the post-COVID assessment environment.** Digital timed exams without proctoring may raise integrity concerns, while untimed digital assessments may be more exposed to AI-assisted completion. Distinguishing genuine performance trends from artefacts of assessment format is a key future priority.

4.5 Policy Implications of UoS (Economics)

The findings highlight several areas that call for institutional action:

- **Assessment design matters for equity.** The temporary closure of gaps during COVID suggests that specific assessment formats can reduce structural barriers. The re-emergence of gaps in written reports and digital untimed exams indicates that longer, open-book tasks may amplify inequalities linked to study environments, independent learning time, or prior educational experience. Institutions should therefore review which formats systematically advantage or disadvantage particular groups.

- **Avoid over-reliance on high-stakes, fully weighted examinations (weight).**Across all groups, post-COVID declines were most pronounced in 100%-weighted final assessments. High-stakes formats concentrate risk and place heavy demands on exam-day performance, time management, and confidence—factors unequally distributed across student groups. Reducing reliance on single, high-stakes tasks and distributing assessment weight more evenly can support fairness for all students while reducing structural disadvantage.
- **Balance equity gains with assessment integrity.** The disappearance of attainment gaps in digital timed assessments post-COVID is encouraging but requires careful interpretation. Un-proctored digital conditions raise the possibility that some improvements reflect masked inequalities rather than the removal of structural barriers. Guidance should therefore balance equity, rigour, and fairness.
- **Generative AI introduces new risks.** Longer open-book assessments are especially vulnerable to AI-assisted completion. The equity implications remain unclear and may widen or mask gaps depending on students' digital literacy and tool access.
- **Diversified assessment portfolios are needed.** No single format works equally well for all groups. Institutions should review how timing (in-term vs out-of-term), format (timed vs untimed), weight (low- vs high-stakes), and delivery mode (digital vs traditional) affect different student groups.
- **Intersectionality should inform student support.** FSM males emerged as a high-risk group in the post-COVID period. Targeted interventions are needed during transition phases (e.g., Level 5 progression).

While methodologically complex, this line of inquiry is essential to ensuring fairness and rigour in evolving assessment practices.

4.6 Economics in Context: Comparison with the Wider Business School

To contextualise the Economics findings, we compare them with results from the wider Business School sample. Although Economics is included within the overall dataset and it represents 31% of the student sample used in the models, this comparison remains meaningful because the Economics cohort represents a distinct subsample within a much larger and more heterogeneous set of programmes.

The purpose is not to contrast fully independent samples, but to assess whether the patterns observed in Economics are distinctive relative to the broader assessment environment. This approach highlights where Economics aligns with wider institutional trends and where it displays discipline-specific dynamics.

As Economics forms part of the full sample, the differences reported here are likely conservative; excluding Economics would make the contrasts with other Business School disciplines even more pronounced. However, several structural features distinguish Economics from the overall Business School population, even though Economics is part of the larger dataset (see Appendix A). When considered on its own, Economics displays the following characteristics:

- a substantially higher proportion of male students (76%)
- lower percentage of students in placement
- greater quantitative intensity, with Economics modules containing roughly 2.6 times more quantitative content than other Business School subjects
- much higher use of digital timed examinations in the post-COVID period (around 50% of all final assessments)

- correspondingly, fewer untimed digital final assessments than Accounting & Finance and Management
- critically, no return to traditional unseen, invigilated exams (UEX), in contrast to Accounting and Finance, which reinstated UEX formats to meet accreditation requirements
- a slightly larger share of fully weighted final exams
- higher rates of mental-health declarations (14% in Economics compared with 8% across the Business School).

These disciplinary characteristics shape both the academic demands placed on students and the assessment environments in which attainment gaps arise, providing essential context for interpreting the findings that follow. Tables 14-16 report some descriptive statistics of the selected department and programmes.

Table 14: Ethnicity at University of Sussex (3 Departments)

BAME	Pre-COVID	COVID	Post-COVID	Total
White	72%	66%	66%	68%
Black	8%	7%	8%	8%
Asian	11%	16%	15%	14%
Mixed-heritage	8%	9%	8%	8%
Other ethnic group	2%	3%	4%	3%
Total	617 (100%)	662 (100%)	832 (100%)	2,111 (100%)
Average number of assessments per student	15	17	16	16
Coursework (In-Term) Final Exam (<100%) Final exam 100% weight	50% 42% 8%	57% 38% 5%	61% 36% 3%	57% 38% 5%
(*) Given the small sample sizes for "Mixed and Other" ethnic minority students, statistical comparisons involving these groups should be interpreted with caution due to limited statistical power.				

Table 15: Distributions of Assessments Across COVID Phases (3 Departments)

Assessment Category	Pre-COVID	COVID	Post-COVID	Total
Short Timed Assessments	25%	28%	37%	31%
Essays	28%	20%	20%	22%
Written Reports	14%	19%	19%	18%
Technical Tasks	9%	9%	5%	7%
Interactive and Engagement-Based Tasks	25%	24%	19%	22%
Total	5,966 100%	9,690 100%	10,545 100%	26,201 100%
Average number of CWK assessments per student	8	10	10	10

Post-COVID coursework assessment across departments shifted towards shorter in term format, with Economics and Accounting and Finance moving furthest in this direction.

Table 16: Final Exams Across Time (3 Departments)

Final Exams	Pre-COVID	COVID	Post-COVID	Legend
Timed Unseen Exam	71%	-	10%	Traditional exams (UEX)
Timed Non -UEX Exam	9%	32%	35%	Including: Open Text Exams (OEX), Seen Exam (EXS), Test (TST) and Computer Based Exam (CEX) and Multiple-Choice Question (MCQ)
Untimed Exam	20%	68%	55%	Including: Essay (ESS), Project (PRG) and Take Away Paper (TAP)
Total	5,847 100%	7,305 100%	6,805 100%	Total final exams 19,957
Average number of final exams per student	7	7	6	6

In terms of assessment format, the three departments differ substantially in the final exam structure:

- Management: ~70% untimed digital assessments (30% timed assessment; no UEX)
- Economics: ~50% untimed and ~50% timed digital assessments (a middle position).
- Accounting & Finance: ~33% untimed, ~28% timed digital, and ~40% traditional UEX (use of UEX due to accreditation requirements).

These structural and assessment-related differences help explain why attainment patterns in Economics may diverge from those observed elsewhere in the Business School.

Awarding gaps: all assessments (3 Departments)

In the wider Business School sample, the Black–White gap did not close during the COVID period. Although performance rose across all groups, Black students remained the lowest-performing group in every phase. This contrasts with the Economics-only sample, where COVID-era assessment changes temporarily reduced inequalities. Therefore, any COVID-related convergence in attainment appears to be discipline-specific rather than institution-wide. This is showed in the figure 5 and test in Table 17.

Figure 5: Ethnic Attainment Trajectories Across COVID Phases (3 Departments)

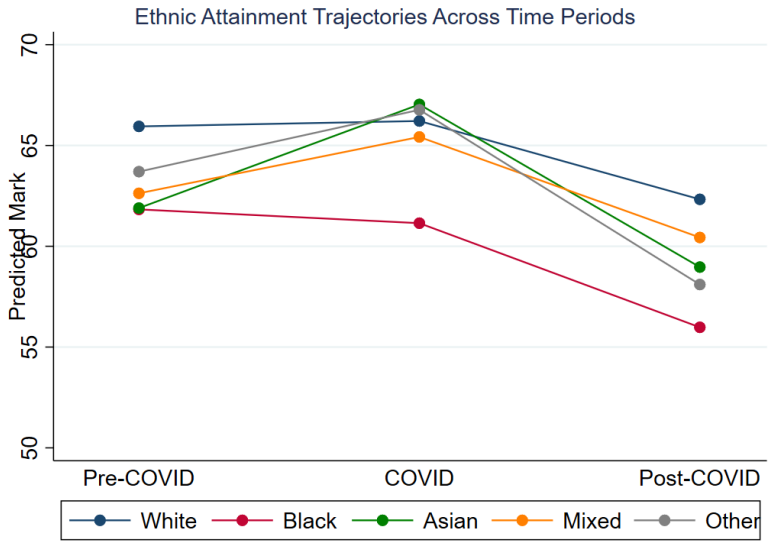


Table 17: Ethnic Awarding Gaps across COVID Phases (overall model) (3 Departments)

Awarding Gaps	Pre-COVID 2018.19	COVID 2020.21	Post-COVID 2022.23	COVID vs. Pre-COVID	Post-COVID vs. Pre-COVID
White vs Black	**4.118*** **	**5.067*** **	**6.349*** **	0.949	2.231
White vs Asian	**4.052*** **	-0.822	**3.361*** **	**-4.874*** **	-0.691
White vs Mixed	**3.319*** **	0.790	**1.893***	**-2.529** **	-1.472
White vs Other	2.244	-0.555	**4.226** **	-2.800	1.982
	Gaps for all main ethnic groups	No gaps except Black	Emergence of gaps for all groups	Asian and Mixed-Heritage benefitted the most	No change between Pre and Post Covid gaps

Note: *** p<0.01, ** p<0.05, * p<0.10.

In the wider Business School sample, awarding gaps during the pandemic followed a different pattern from the one observed in Economics. Although all ethnic groups displayed significant gaps relative to White students before COVID in both datasets, the pandemic period produced divergent effects across the two samples.

- During COVID:
  - Economics: all awarding gaps disappeared, including the Black–White gap.
  - Business School (overall):
    - Gaps disappeared for Asian, Mixed-Heritage, and Other groups.
    - Black–White gap persisted, showing no COVID-era convergence.
- After COVID:
  - Business School: awarding gaps re-emerged for all ethnic groups.
  - Economics: the widening was focused mainly on Black students, not all groups.

As in Economics, however, the reappearance of gaps occurred alongside a general decline in average performance across all groups, suggesting that post-COVID cohorts faced broader academic challenges regardless of ethnicity.

Awarding gaps: in-term Coursework (3 Departments)

Figure 6 and Table 18 show that across the wider Business School sample, awarding gaps follow the same broad temporal pattern. Before the pandemic, significant gaps were present for almost all ethnic groups in several assessment types, particularly in essays. During COVID, gaps disappeared for all groups except Black students, who remained the lowest-performing group across most assessment types. Post-COVID, awarding gaps re-emerged for all ethnic groups, with the largest disadvantages appearing across nearly all assessment formats (except technical reports).

Figure 6: Ethnic Attainment in Coursework (3 Departments)

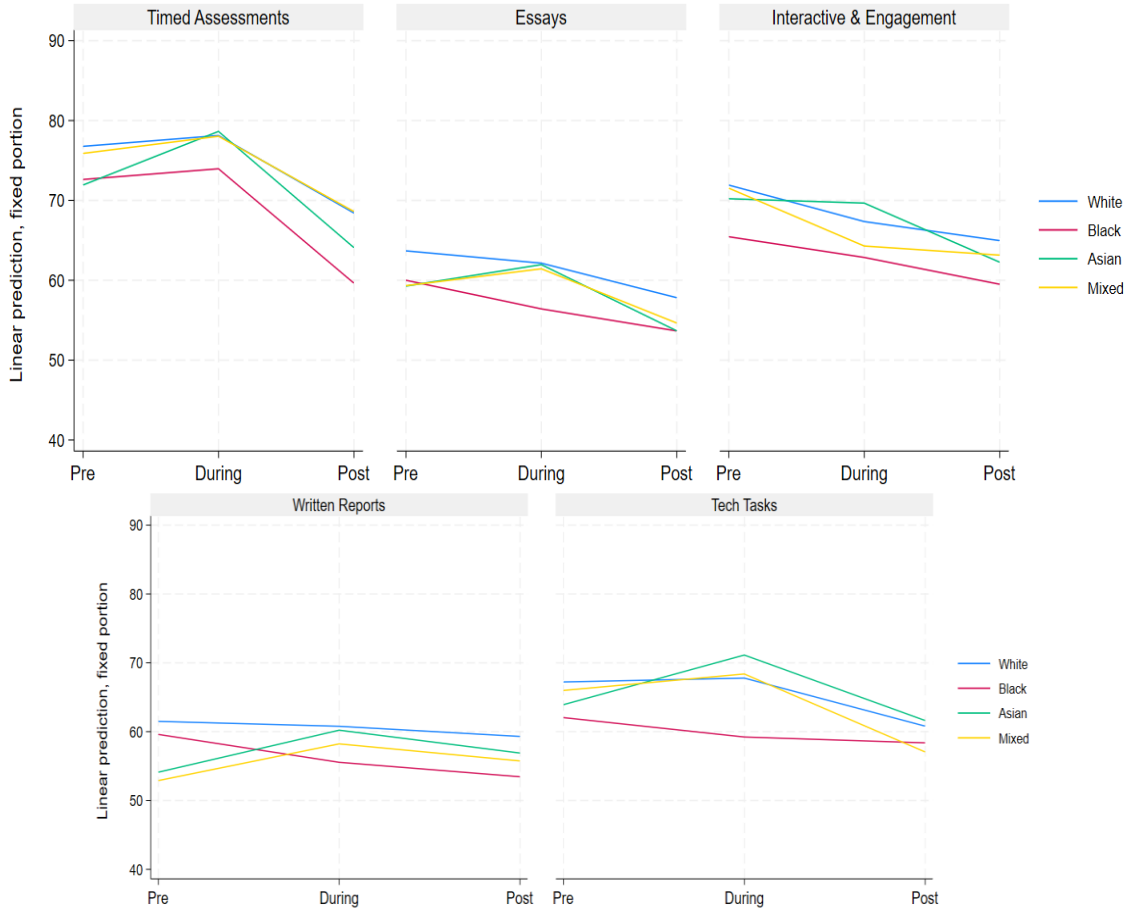


Table 18: Post-COVID Awarding Gap in Coursework (3 Departments)

Ethnicity comparison	Short Timed Assessments	Essays	Written Reports	Technical Tasks	Interact. and engagement Task
White vs Black	**8.783*** **	**4.161** **	**5.859*** **	2.430	**5.463** **
White vs Asian	**4.336*** **	**4.154*** **	**2.415** **	-0.816	2.702
White vs Mixed	-0.189	**3.171** **	**3.561** **	3.745	1.816
White vs Other	3.88	3.032	**6.822** **	-0.287	4.453
Result	Gaps: Black and Asian groups	All groups (except "Other"	**Gaps: all ethnic groups**	No gaps	Gap : only black students
Note: * p<0.10, ** p<0.05, *** p<0.01					

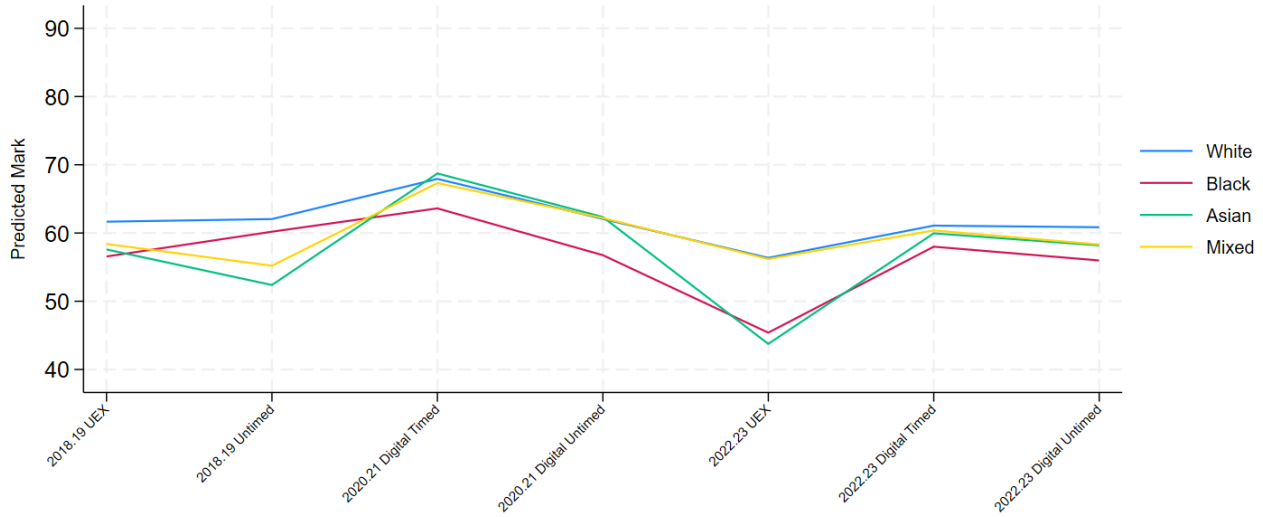
By contrast, in the Economics subsample, all coursework gaps disappeared during COVID. Post-COVID, gaps re-emerged only for Black students and were concentrated mainly in written reports. These patterns suggest that the COVID-phase equalisation observed in Economics was discipline-specific and becomes diluted when combined with the more heterogeneous assessment practices of the wider Business School.

Awarding gaps: Final Exams (3 Departments)

Figure 7 and Table 19 reveal the following: across the wider Business School sample, final-exam awarding gaps follow a clear and consistent pattern.

- Before the pandemic, significant gaps were present for all ethnic groups across most assessment formats, with the only exception being timed non-UEX exams. (By comparison, in the Economics-only subsample, pre-pandemic gaps were concentrated primarily in UEX exams for Black students.)
- During COVID, the shift to digital assessment—unlike in Economics—did not eliminate gaps: Black students remained disadvantaged in both digital timed and digital untimed formats.
- In the post-COVID period, again unlike Economics, awarding gaps re-emerged for all ethnic groups, especially in UEX and digital untimed exams, while digital timed exams continued to show no significant disparities.

Figure 7: Ethnicity Attainment in Final Exams (3 Departments)



Note: Predicted margins controlling for student and module characteristics.

Table 19: Ethnicity Attainment in Final Exams (3 Departments)

Ethnic Group	2018-19 UEX	2018-19 Timed exams	2018-19 Untimed exams	2020-21 Timed exams (digital)	2020-21 Untimed exams (digital)	2022.23 UEX	2022-23 Timed exams (digital)	2022-23 Untimed exams (digital)
White vs. Black	**5.098*** **	1.209	1.849	**4.324** **	**5.31*** **	**10.91*** **	3.091	**4.871*** **
	(p=0.001)	(p=0.722)	(p=0.252)	(p=0.038)	(p=0.002)	(p=0.002)	(p=0.126)	(p=0.008)
White vs. Asian	**4.066** **	-1.269	**9.663*** **	-0.791	-0.282	**12.61*** **	1.131	**2.655** **
	(p=0.023)	(p=0.619)	(p=0.002)	(p=0.523)	(p=0.782)	(p=0.000)	(p=0.362)	(p=0.042)
White vs. Mixed	**3.257** **	-0.579	**6.838** **	0.614	-0.134	0.175	0.713	**2.523** **
	(p=0.027)	(p=0.895)	(p=0.015)	(p=0.663)	(p=0.907)	(p=0.969)	(p=0.633)	(p=0.045)
White vs. Other	3.951	-4.515	13.871	-1.512	-0.473	3.839	2.743	**3.615***
	(p=0.237)	(p=0.500)	(p=0.118)	(p=0.349)	(p=0.755)	(p=0.197)	(p=0.356)	(p=0.056)
Result: Gaps group	Gaps (except Other)	No gaps	Asian and Mixed Gaps	**No gaps except Black students**		Gaps Black and Asian	No gaps	**All gaps!**

Note: * p<0.10, ** p<0.05, *** p<0.01.

A further difference from in-term coursework is that post-COVID performance in final exams did not fall below pre-COVID levels. After the temporary uplift during the COVID year, final-exam marks returned broadly to their pre-pandemic levels.

The absence of awarding gaps in digital timed exams raises important questions about interpretation. Digital timed formats may genuinely reduce structural barriers by limiting subjective marking variation and reducing reliance on extended writing. However, in un-proctored online settings, they may also risk masking inequalities through unauthorised collaboration or external resource use.

Importantly, timed exams also showed no gaps before the pandemic—although they represented only about 9% of all pre-COVID final assessments and were predominantly non-digital. This suggests that certain intrinsic features (e.g., structured questions or MCQ-based recall) may support more equitable performance across groups. At the same time, because timed exams accounted for a much larger share of post-COVID assessments (around 35%), further evidence is needed to determine whether the observed equity reflects genuine removal of structural barriers or shifts in assessment integrity.

Our last observation is about socioeconomic variable and performance (Table 20). FSM patterns in the wider Business School sample are more straightforward than in Economics. FSM students showed a clear disadvantage before the pandemic, but this penalty disappeared during COVID and did not re-emerge afterwards across any of the three models (overall, coursework, or final exams). The reasons for this sustained equalisation, despite the small increase in FSM students, are difficult to isolate, as several factors changed simultaneously—including assessment format, support structures, and temporary adjustments to entry requirements.

By contrast, in Economics, FSM patterns were more complex and interacted with gender: FSM penalties were small or absent pre-COVID, converged during COVID, and re-emerged specifically for FSM males in the post-COVID period. No comparable post-COVID penalty was observed for FSM females. Economics presents a combination of structural and pedagogical features that may amplify challenges for some students. In particular, the male-dominated culture, the high quantitative modules alongside analytical writing requirements, heavier use of high-stakes final examinations, and the prevalence of timed digital examinations may create pressures that could disproportionately affect FSM male students.

Table 20: FSM Disadvantage by Time Period (3 Departments)

FSM EFFECT IN EACH PERIOD	**Model 1** (Overall)		**Model 2** (Coursework)		**Model 3** (Final Exams)	
	Coeff.	p-value	Coeff.	p-value	Coeff.	p-vale
Pre-COVID (2018.19)	**-5.753*** **	0.003	**-4.023** **	0.026	**-7.721*** **	0.001
COVID (2020.21)	-0.607	0.604	-1.480	0.226	-0.636	0.576
Post-COVID (2022.22)	-1.136	0.427	-1.553	0.281	-1.902	0.227

Note: * p<0.10, ** p<0.05, *** p<0.01; Negative coefficients indicate FSM disadvantage (FSM students scored lower than non-FSM peers).

Finally, the effects of controlling factors are consistent and behave similarly across models (female advantage, disability effects, Level 6 progression). The results of the three models for

the Business School are reported in Appendix A.3. When we compare them with Economics results, several critical differences emerge:

-First, Level 5 progression shows a striking department-specific divergence. In the broader Business School, Level 5 students achieve substantial gains relative to Level 4 and continue at Level 6. In Economics, no such advantage appears, indicating a unique mid-degree challenge specific to Economics.

-Second, the FSM disadvantage disappears post-COVID across the full Business School sample, and no gender intersectionality is at play.

-Third, ethnic-gap patterns differ meaningfully between Economics and the wider School. In the full Business School sample, post-COVID gaps re-emerge earlier and across a broader set of assessment types, whereas in Economics the largest disparities are concentrated specifically in extended written assessments and untimed digital finals.

-Fourth, quantitative module effects follow a consistent pattern—positive in coursework but negative in final exams—yet Economics' 2.6× higher quantitative intensity means that far more students are exposed to the exam-based quantitative penalty. This magnifies the impact of quantitative assessment in Economics relative to other departments.

-Finally, post-COVID declines in coursework performance are substantially larger in the broader Business School than in Economics alone.

4.7 Conclusions of UoS (Business School: 3 Departments)

The analysis of the wider Business School sample shows that ethnic attainment gaps are persistent but vary across time, assessment type, and format. Three central conclusions emerge:

1. **Pre-existing awarding gaps were substantial and widespread.** Before the pandemic, almost all ethnic groups showed significant disadvantages in several final-exam and coursework formats, particularly essays and untimed assessments.
2. **COVID-19 reduced—but did not eliminate—inequalities.** The move to digital assessment narrowed attainment gaps, but unlike Economics, convergence was incomplete. Black students remained consistently disadvantaged across most formats, including digital assessments.
3. **Post-COVID, gaps re-emerged across all groups.** After returning to stable assessment practices, all ethnic groups experienced renewed disadvantages in most assessment types. The largest post-COVID gaps arose in written reports, short-timed assessments, and untimed final examinations. In contrast, digital timed exams continued to show no significant gaps.

Beyond ethnicity, FSM disadvantage—strong and significant pre-COVID—disappeared during COVID and remained statistically insignificant afterwards. This contrasts with Economics, where FSM disadvantage re-emerged for FSM males' post-COVID. The Business School therefore appears less affected by intersectional socioeconomic patterns than Economics.

Final-exam performance returned to pre-pandemic levels across the School, while coursework performance fell markedly, suggesting different recovery trajectories and assessment sensitivities.

4.8 Limitations of UoS analysis (Business School: 3 Departments)

Several caveats should be noted:

- **Economics is part of the School dataset**, so comparisons between "Economics" and "Business School" are conservative; excluding Economics would likely amplify differences.
- **Some assessment formats have small subgroup samples**, especially UEX exams or "Other" ethnic groups, limiting statistical power.
- **The analysis is based on summative results only.**
  Day-to-day academic experience, feedback processes, and learning behaviours are not observed here. (Though these are explored in the project's focus group strand.)
- **FSM patterns are difficult to interpret confidently.**
  Even though FSM disadvantage disappears post-COVID, we cannot isolate whether this is due to assessment flexibility, cohort characteristics, policy shifts, or support structures.
- **Integrity concerns in digital timed assessments remain unresolved.**
  The consistent absence of awarding gaps in timed digital exams could reflect both genuine equity gains and potential masking effects from un-proctored conditions.

4.9 Policy Implications for UoS (Business School: 3 Departments)

The Business School results suggest several areas where institutional reforms may improve equity:

1. **Assessment design remains a powerful driver of inequality.** Gaps are smallest in structured, time-limited tasks and largest in extended, open-ended assessments (essays, written reports, untimed exams). Programme leaders should review the balance of assessment types to ensure that no single format disproportionately drives disadvantage.

2. **Distinguish genuine equity from potential masking in timed digital exams.** Gaps disappear in timed digital exams, but this could reflect both reduced bias and risks to assessment integrity. Guidance is needed to balance rigour, authenticity, and equity when using un-proctored digital formats.

3. **Written reports and untimed exams require redesign.** These formats show the largest awarding gaps post-COVID. They rely heavily on independent study space, time availability, and extended writing proficiency—factors unevenly distributed across student groups. Scaffolding, alternative formats, or redesigned marking criteria may mitigate disparities.

4. **Maintain diversified assessment portfolios.** No single assessment format is "fair for all." Departments should map awarding gaps by assessment type and ensure that students are not disproportionately exposed to formats associated with larger inequalities.

5. **Socioeconomic disadvantage appears mitigable.** FSM penalties disappeared during and after COVID. This suggests that supportive structures (flexible assessment, improved access, or additional learning resources) can eliminate socioeconomic gaps. These practices should be retained and embedded.

6. **Monitor Level 5 and Level 6 differentially.** Progression patterns indicate that the mid-degree (Level 5) is a point of increased difficulty across the School. Targeted

academic support during this year may prevent widening gaps. Having examined both the Economics programme and the wider Business School context at UoS, we now turn to the analysis QMUL. The same modelling strategy is applied, enabling meaningful comparison across institutions while taking account of their differing programme structures and assessment practices. Presenting the UoS and QMUL findings side by side allows us to assess whether patterns in assessment and awarding gaps are consistent across institutions or shaped by institution-specific factors.

5. Chapter 3: Results for Queen Mary University (QMUL)

5.1 Reference Categories and Institutional Context

The QMUL regression models used Asian students and the During-COVID period (2020–21) as reference categories, in contrast to Sussex, where the baseline was White students and the Pre-COVID period. These choices reflect institutional differences in student composition and assessment timelines. At QMUL returned to traditional, in-person exams in 2022–23, making the COVID year—with its emergency shift to online assessment—a natural midpoint for evaluating change. Additionally, at QMUL, Asian students constituted the majority of the sample (approximately 68%, $n \approx 340$), making them a statistically efficient reference group. However, since policy discussions centre on differences relative to White students, we reframe our output [10] presenting all group differences with White students as the reference (Tables 21-24).

The full regression outputs from all three models — overall performance, in term assessments (in-term), and final exams (end-of-term) — are reported in the Appendix B. This section summarises the main findings, with particular attention to how ethnicity interacted with time period and assessment type. All estimates control for student demographics, module characteristics, and institutional variables, as described in the modelling strategy.

5.2 All Assessment Types: Overall Model (Economics)

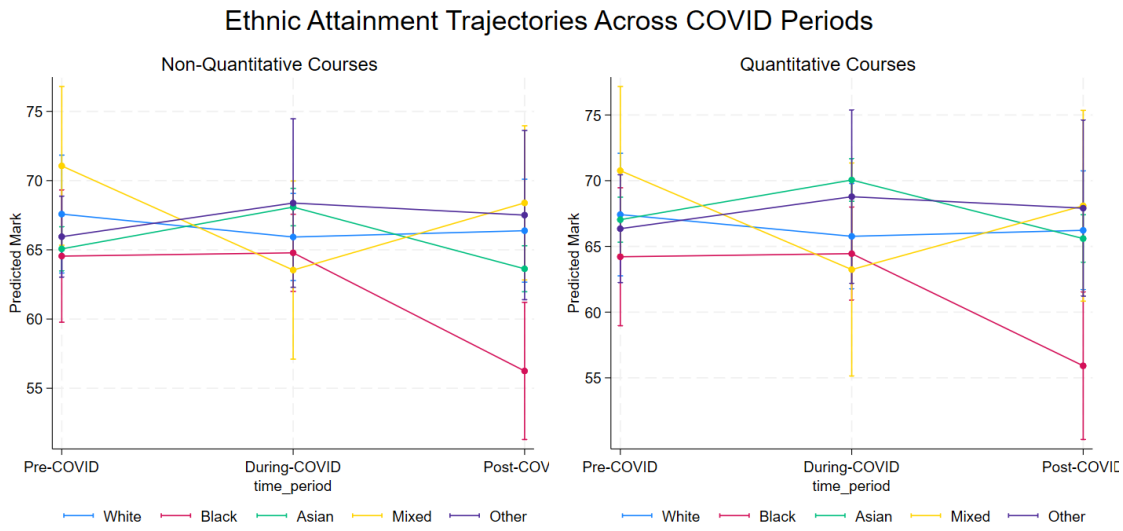
Figure 8 presents predicted performance values (i.e. marginal effects at representative values, with categorical controls at their reference categories and continuous controls set at sample means) for all groups across the three time periods, separately for non-quantitative and quantitative. Figure 9 is the same as Figure 8, but with a focus on Asian, White, and Black students.[11] The Figures highlight the main visual patterns, which are then formally tested, in Tables 21–22. In summary: .

- Asian students show a peak during the COVID online assessment period, followed by a return to pre-COVID levels post-COVID (return to in-person exams).
- Black students show the sharpest deterioration, with a marked decline in the post-COVID period.
- White students remain broadly stable and have the highest performance post-COVID.
- Awarding gap (vs. White group) convergence during COVID and a large divergence post-COVID, driven by a widening White–Black gap.
- Across modules, Asian–White differences narrow more in quantitative modules (reflecting Asian students' quantitative advantage), whereas Black students experience severe post-COVID losses in both module types.

[10] We used the command margins, for the predicated value and post-estimation linear combinations (*lincom* command in Stata) for testing gaps within and across groups.

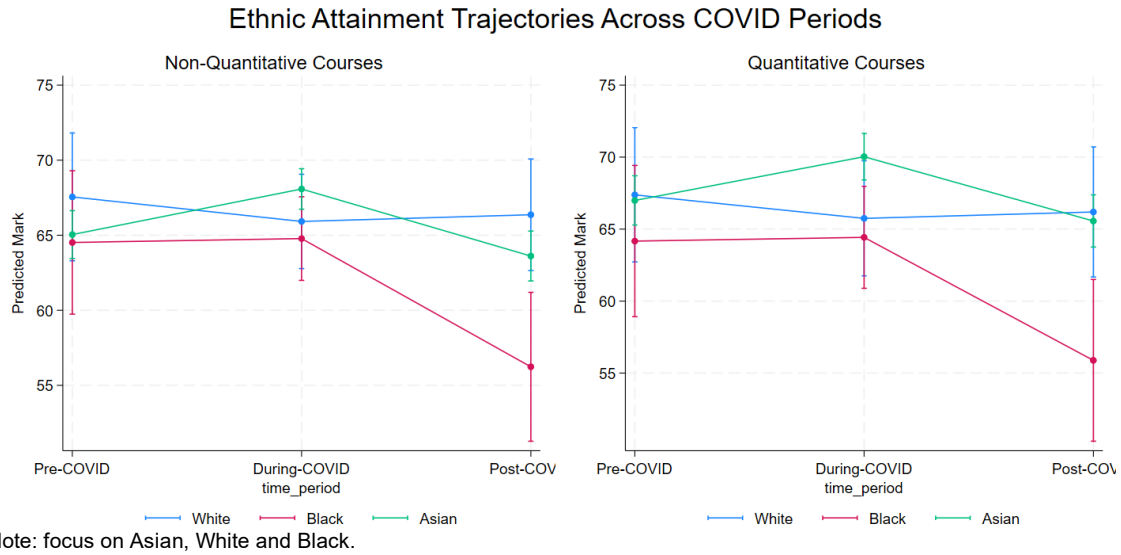
[11] Estimates for Mixed and Other ethnic groups (each representing 4% of the sample) should be interpreted with caution due to small sample sizes and limited statistical power. We excluded then from the Tables 21-22 to focus on the three main groups (Asian, White and Black students) due to small sample sizes. Results are reported in the Compendium (Tables C.7-C.12).

Figure 8: Ethnic Attainment Trajectories Across COVID Phases (Economics)



Note: All Ethnicity groups.

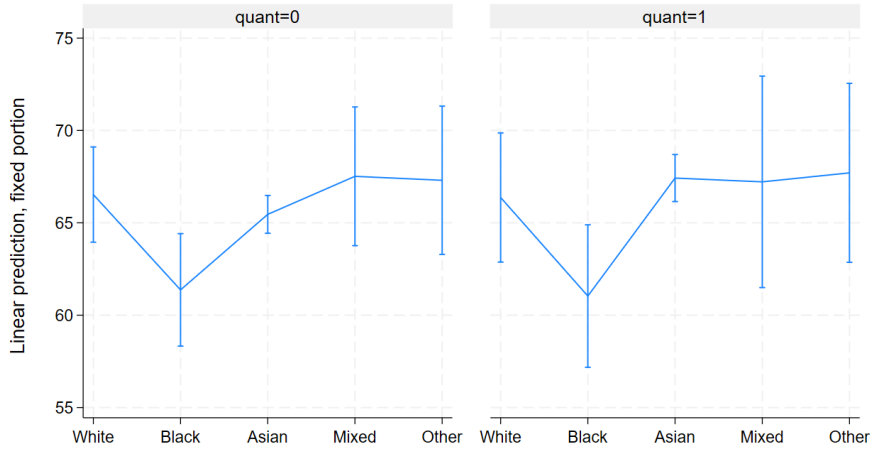
Figure 9: Ethnic Attainment Trajectories Across COVID Phases (Economics)



Note: focus on Asian, White and Black.

Asian students show a rise during the emergency online period, exceeding both their pre-COVID and post-COVID levels, particularly in quantitative modules (Figure 10).

Figure 10: Ethnic Attainment in Non-Quantitative and Quantitative Modules (Economics)



Note: Vertical bars represent 95% confidence intervals.

Table 21 confirms that this temporary online advantage, and their within-group quantitative strength, is statistically significant. Once in-person examinations resumed, Asian performance returned to approximately pre-pandemic levels but retained a within group advantage in quantitative versus non quantitative modules, relative to the White and Black groups. This latter group shows a marked post-COVID deterioration, well below both their pre-COVID and COVID levels. Table 21 indicates that these declines are large in magnitude and statistically robust. White students show no meaningful change across time or module type, which is consistent with the flat pattern visible in both panels.

Table 21: Within Group Differences (Asian, White and Black students) (Economics)

Ethnicity	Academic year			Module
	Pre-Covid vs. Covid	Post-Covid vs. Covid	Post-Covid vs. Pre-Covid	Quant vs. Non-Quant
White	2.267	0.823	-1.444	-0.680
Black	-0.376	**-8.165*** **	**-8.541*** **	-0.856
Asian	**-2.405** **	**-4.088*** **	-1.683	**1.443*** **

Note: *** p<0.01, ** p<0.05, * p<0.10. Time columns show predicted change in marks for each ethnic group relative to their COVID (2020-21) baseline. Module column shows quantitative module effects (difference between quant and non-quant performance) for each group.

These group-specific trajectories can explain the cross-group emergence of a large White–Black gap after the COVID period, during the post-COVID phase, as reported in Table 22. The post-pandemic gap is statistically significant in both quantitative and non-quantitative modules, whereas it was not present before COVID. Importantly, the widening occurs after assessment returned almost entirely to the more conventional pre-Covid formats, suggesting that the inequality emerged in the recovery phase rather than during the emergency online period itself.

Table 22: Ethnic Awarding Gaps Across COVID Phases (Economics)

Ethnicity	Ethnicity attainment Gaps		
	Pre-Covid	Covid	Post-Covid
White vs. Black (quant=0)	3.034 (p=0.355)	1.142 (p=0.593)	**10.130*** ** (p=0.001)
White vs Asian (quant=0)	2.511 (p=0.281)	-2.1612 (p=0.216)	2.749 (p=0.172)
White vs Black (quant=1)	3.210 (p=0.371)	1.319 (p=0.629)	**10.308*** ** (p=0.005)
White vs Asian (quant=1)	0.388 (p=0.879)	**-4.284*** (p=0.053)	0.627 (p=0.797)
Result	No gaps	Advantage gap for Asian in Quant	Gaps for Black in quant and in non-quant

Note: *** p<0.01, ** p<0.05, * p<0.10.

The White–Asian comparison shows a different mechanism. Asian students' quantitative strength reduces the White–Asian gap, a pattern confirmed by Table 23. However, overall post-COVID differences between Asian and White students are not statistically significant, suggesting that the COVID-period gains for Asian students were transient and module-specific rather than structural.

Table 23: Ethnic Awarding Gaps in Quantitative and Non-Quantitative modules (Economics)

Ethnicity	Non-quant Gap	Quant Gap	Gap change
White vs. Black	1.142 (p=0.593)	1.319 (p=0.629)	0.176 (p=0.908)
White vs. Asian	-2.162 (p=0.216)	**-4.284*** (p=0.053)	**-2.112**** (p=0.041)

Note: *** p<0.01, ** p<0.05, * p<0.10. Gaps shown are average across all time periods. Positive values in columns 1-2 indicate White advantage. Positive values in column 3 (Gap change) indicate gap is larger in quantitative modules.

Taken together, these results indicate that COVID-related attainment effects were heterogeneous. The critical equity concern is the emergence of a substantial White–Black gap that persists across module types and assessment formats, indicating that Black students faced broader recovery-related disadvantage that cannot be mitigated simply through changes in curriculum composition or assessment modality.

Additional model effects, common to all ethnic groups, are reported in Table 24. Female students show a modest overall advantage (+1.8 points, p=0.043), which is larger in quantitative modules (+1.6 points, p=0.024), indicating stronger relative performance in quantitative content. Students with disabilities display a reversal: they perform higher at the COVID baseline (+4.3 points, p=0.013) but substantially lower pre-COVID (–9.0 points, p=0.002), suggesting that emergency online assessment conditions were particularly beneficial for this group.

Students performed better in core than optional modules (1.9 points, p<0.001), suggesting that standardized curriculum and pedagogical investment may matter more than student

choice. Performance also improved with progression: both level 2 and level 3 modules scored approximately 4 points higher than level 1 ($p<0.001$), indicating successful academic development across the degree.

Students from higher professional backgrounds achieve higher marks overall (+3.0 points, $p=0.008$), consistent with persistent socioeconomic advantage. Out-of-term assessments (i.e. the final examinations; online during COVID and in-person outside the COVID period) show a small positive baseline difference (+0.25 points, $p=0.001$), but this advantage diminishes pre- and post-COVID (interactions $p<0.01$), indicating that the benefit was specific to the emergency online context rather than sustained under normal in-person examination conditions.

Table 24: Additional Model Effects - Control Variables and Interactions (Economics)

Variables	Coefficient	SE
A. ASSESSMENT & MODULE EFFECTS		
Out-of-term timing (main effect)	**0.251*** **	0.074
× Pre-COVID	**-0.421*** **	0.120
× Post-COVID	**-0.365** **	0.117
Core module	**1.894*** **	0.363
Term 2 vs Term 1 assessment	**-0.622***	0.260
Module level 2	**4.038*** **	0.900
Module level 3	**3.680*** **	0.740
B. GENDER EFFECTS		
Female (main effect)	**1.835***	0.907
Female × Quantitative	**1.631***	0.723
C. DISABILITY CONDITION EFFECTS		
Any disability (COVID baseline)	**4.278***	1.730
× Pre-COVID	**-9.012** **	2.894
× Post-COVID	-4.547	3.250
D. SOCIOECONOMIC EFFECT		
High professional occupation	**2.978** **	1.123
E. RANDOM EFFECTS (SD)		
Student intercept	8.055	0.616
Core module slope	6.487	0.378
Module level slopes	7.486-8.915	0.622-0.761
Residual	7.345	0.129
Note: *** p<0.001, ** p<0.01, * p<0.05. Reference: COVID period, male, no disability, non-professional background, in-term, non-quantitative, non-core modules, module level 1.		

5.3 In-Term Assessments: Coursework Model (Economics)

We restrict now to coursework only (i.e., different types of coursework occurred during the teaching term) and we consider different types of assessments. Within this subset, four coursework formats were distinguished to capture to capture differences in task structure: (i) short timed coursework (tests, MCQs, quizzes, brief tasks, problem sets, homework, portfolio-

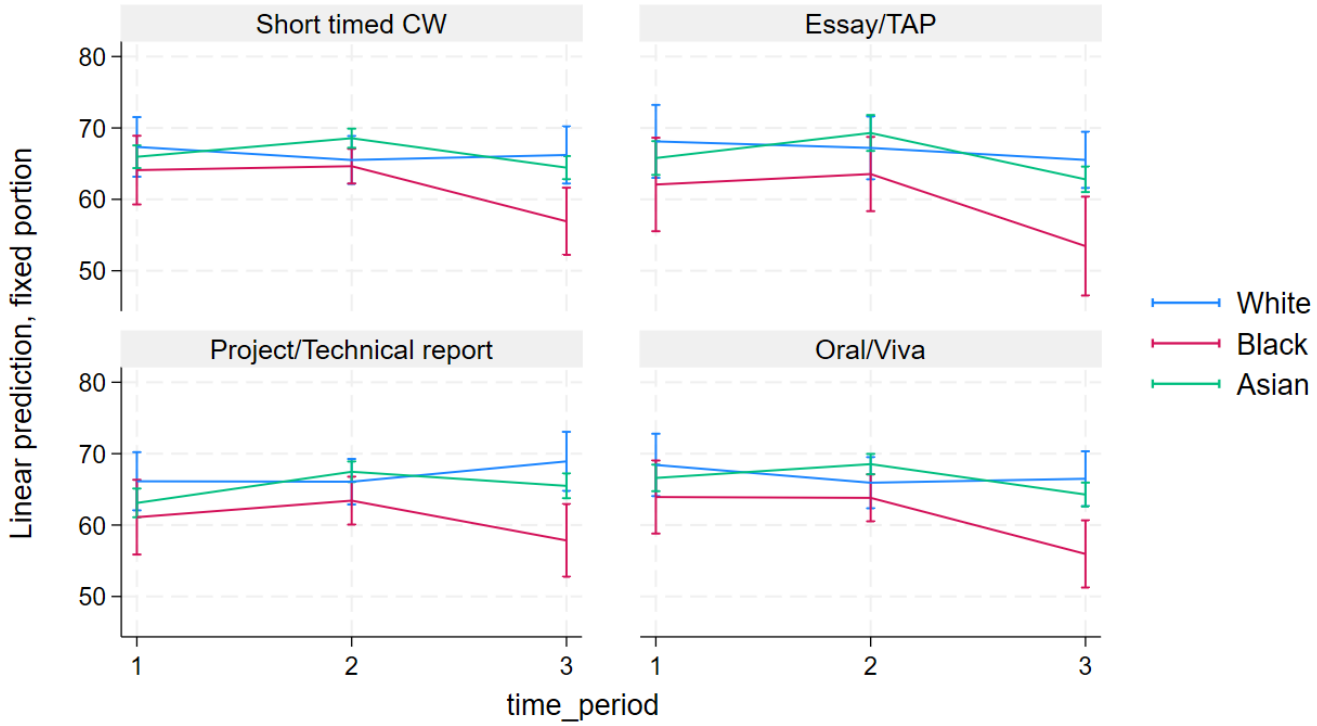
type submissions); (ii) extended written coursework (essays / TAP); (iii) technical / project-based coursework (projects, technical reports, poster-type outputs); and (iv) oral coursework (presentations, vivas). This classification allows us to test whether the post-COVID divergence in attainment is specific to certain coursework formats, or whether the same dynamic is visible across all of them.

Because assessment formats changed over time (with certain types concentrated during COVID and others post-COVID), period effects and assessment-type effects are not fully independent. As such, observed period differences partly reflect both timing and institutional compositional changes in assessment formats.

Restricting the sample to in-term coursework did not change the predicted period-specific ethnic margins, and the predicted values by ethnicity and time period were almost identical to those obtained in the full assessment set. Asian students again show a rise-and-return trajectory, whereas the sharp post-COVID drop for Black students persists. Given the practical identity of these period margins to those from the full model, we do not re-present them here and focus instead on variation across coursework types.

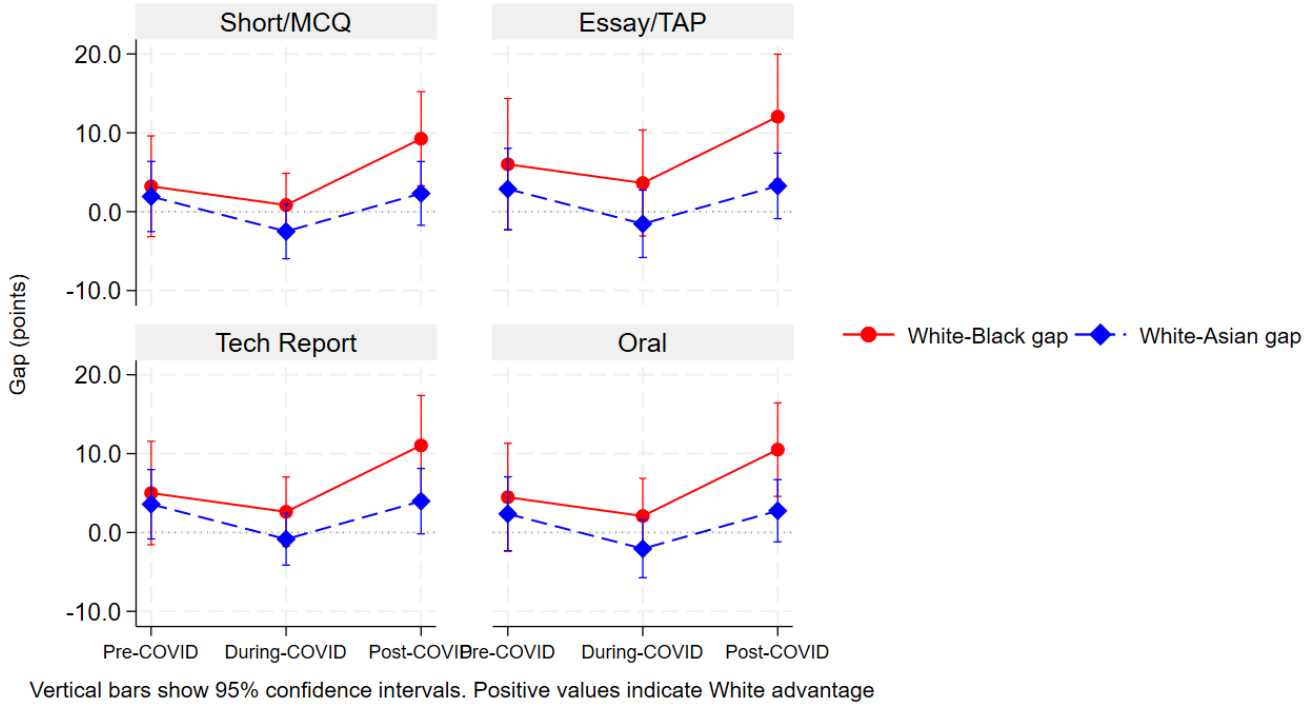
The coursework-specific margins (Figure 11 for predictive margins and Figure 12 for gap version of predictive margins) show that the post-COVID widening of the White–Black gap is visible in every coursework format, while White–Asian differences remain small and comparatively stable. In other words, the emergence of the White–Black gap is not driven by a particular coursework mode.

Figure 11: Ethnic Attainment by In-term Assessments Across COVID Phases (Economics)



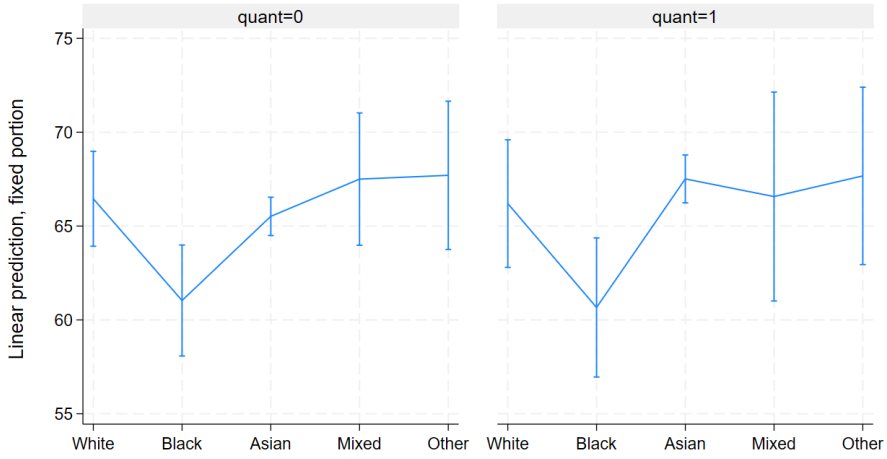
Note: time period (1 = Pre-COVID, 2=COVID, 3=Post-COVID). Vertical bars show 95% confidence intervals.

Figure 12: Ethnic Attainment Gaps Across COVID Phases, In-term Assessments (Economics)



The quantitative interaction provides an important mechanism that helps explain the stability of White–Asian gaps and the deterioration of White–Black gaps. Figure 13 reports the predicted grades for each ethnic group, separately for non-quantitative and quantitative modules.

Figure 13: Predictive Margins by Non-Quantitative and Quantitative Modules, In-term Assessments (Economics)



The within-group estimates show that Asian students obtain markedly higher marks in quantitative modules than in non-quantitative modules. The opposite is true for White and Black students who do not show a comparable "quantitative bonus and for both groups'

performance falls in quantitative formats, with the decline for Black students slightly larger than for White students.[12]

This asymmetry matters for the gap. Asian students' quantitative strength acts as a protective mechanism: because Asian students constitute the largest ethnic group in the sample, and quantitative coursework represents a meaningful share of term-time assessment (around one-quarter of tasks), this within-group advantage carries enough weight to prevent a large White–Asian separation from emerging. By contrast, Black students incur a quantitative penalty relative to their own non-quantitative performance, so quantitative coursework does not improve their relative position and can reinforce their disadvantage.

This explains why White–Asian differences remain small and stable across periods, whereas the post-COVID gap opens almost entirely between White and Black students. In short, quantitative formats help buffer Asian outcomes, but they do not contribute much to generate the White–Black divergence. The widening White–Black gap reflects a broader post-pandemic disadvantage for Black students rather than the composition of coursework.
Restricting the sample to coursework formats also confirms that assessment mode does not explain the divergence: White and Asian students remain closely aligned across all coursework types, whereas Black students experience a sharp post-COVID drop in every format. Finally, the coefficients in the coursework-only model are substantively indistinguishable from the full model, so the interpretation of the non-ethnicity controls remains unchanged.

Finally, the coefficient estimates from the coursework-only model are substantively indistinguishable from the full model (reported in Table 26 at the end of the section). Gender, disability, and socioeconomic effects remained similar in magnitude and direction, and no new interactions emerged. Given the consistency of these estimates, we do not reproduce the full table here and focus instead on the coursework assessment-type specification, where the final inferential tests are introduced.

5.4 Out-of-term Assessments: Final Exams Model (Economics)

Because examination delivery shifted from in-person pre-pandemic to fully online during COVID and then largely returned to in-person afterwards, final examinations were grouped into four delivery–format categories:

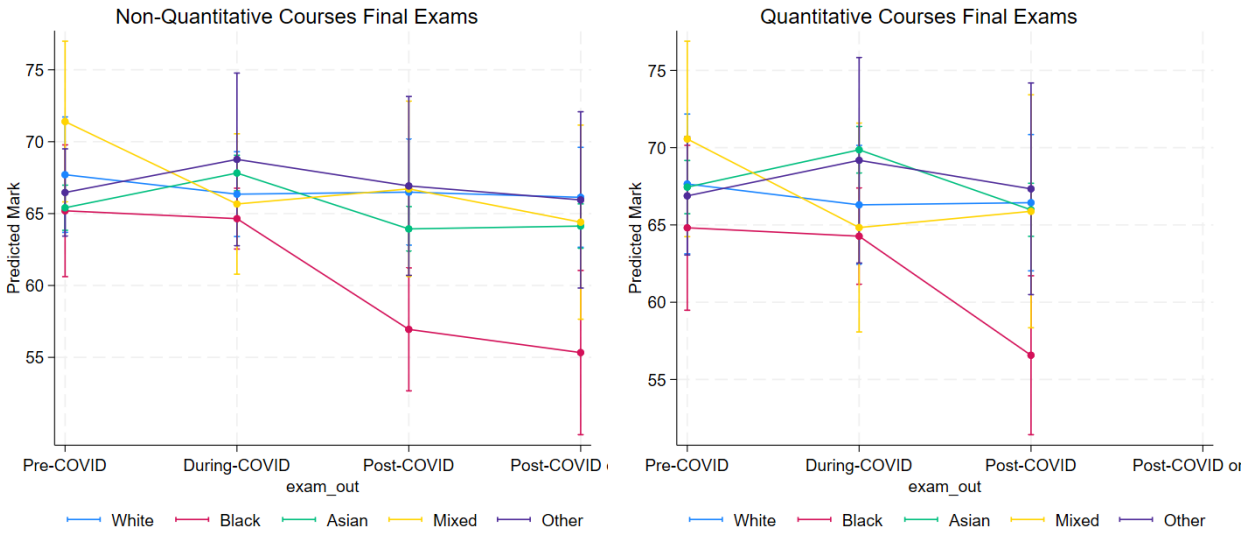
1. Pre-COVID in-person exams,
2. COVID emergency online exams,
3. Post-COVID in-person exams, and
4. Post-COVID online-style assessments (e.g., reports, TAPs).

This classification allows us to distinguish changes linked to delivery mode from those arising during the broader recovery phase.

Across final exams, non-quantitative modules accounted for roughly three-quarters of all assessments (about 75%), while quantitative modules made up the remaining 25%; in the post-COVID period, non-quantitative modules continued to use both timed and untimed digital formats, whereas quantitative modules reverted entirely to a single in-person style of assessment, which is why Figure 11 shows only one post-COVID point for quantitative modules.

[12] In the figure with gaps, if the vertical line does not cross zero, it means that the gap is statistically significant at 5%.

Figure 14: Ethnic Attainment in Final Exams (Economics)



Note: Vertical bars show 95% confidence intervals.

Across all formats, ethnic gaps were small and statistically non-significant in both the pre-COVID and COVID phases (Table 25). During the COVID year, White and Black students performed at similar levels under the emergency online assessment regime. The only clear exception was Asian students, who achieved their highest marks—especially in quantitative modules (around 70–71 points)—indicating that the online environment aligned well with their comparative strengths.

Table 25: Awarding Gaps by Type of Exams and Modules (Economics)

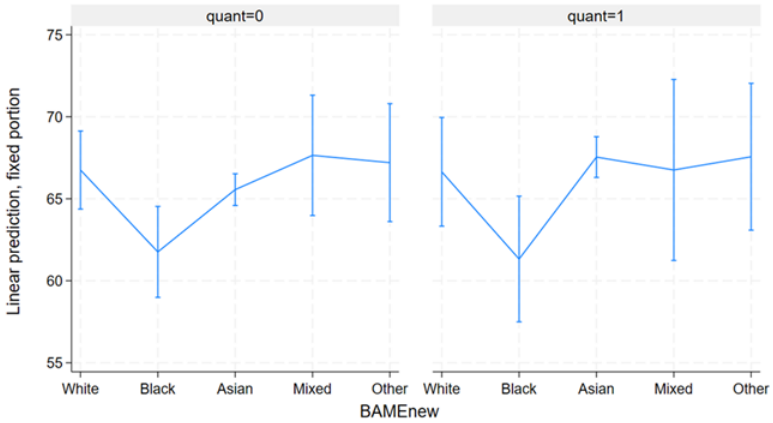
Exam Types	White-Black	p-value	White-Asian	p-value	White-Mixed	p-value
NON-QUANTITATIVE MODULES						
Traditional UEX (Pre-COVID)	2.51	0.420	2.30	0.297	-3.70	0.294
COVID Digital	1.70	0.361	-1.47	0.371	0.69	0.814
UEX (Post-COVID)	**9.55*** **	0.001	2.56	0.202	-0.23	0.951
Untimed (Post-COVID)	**10.8*** **	0.002	2.00	0.298	1.73	0.657
QUANTITATIVE MODULES						
Traditional UEX (Pre-COVID)	2.83	0.425	0.21	0.934	-2.91	0.465
COVID Digital	2.03	0.423	**-3.56***	0.093	1.47	0.712
UEX Post-COVID	**9.87*** **	0.004	0.46	0.847	0.56	0.901

A large and statistically significant White–Black gap emerges only in the post-COVID period. This occurs across all final-exam formats: in-person invigilated exams and online-style out-of-term assessments. The gap appears even though post-COVID exams largely reverted to their pre-pandemic format, when no White–Black gap had been present.

Figure 15 highlights an important feature:

1) Asian students benefit from quantitative exams relative to their own non-quantitative performance.
2) White students perform slightly worse in quantitative modules.
3) Black students experience a decline in both quantitative and non-quantitative formats.

Figure 15: Ethnic Attainment in Final Exams of Non-Quantitative and Quantitative Modules (Economics)



Note: Vertical bars show 95% confidence intervals. Non-quantitative module=0 and quantitative=1.

Taken together, these patterns are consistent across coursework, final exams, and both quantitative and non-quantitative contexts. The two post-COVID data points—representing in-person exams and flexible online-style assessments—cluster closely, with overlapping confidence intervals. The mechanism driving these patterns becomes clear through this comprehensive view: Asian students' selective quantitative advantage helps to avoid gaps, while the widening White-Black gap stems from a generalized deterioration in Black students' performance across all contexts.

This indicates that the mechanism underlying the widening ethnic gaps cannot be attributed solely to exam modality. Pre-COVID in-person exams did not produce a substantial White–Black gap, and during the COVID year fully online assessments also showed no divergence between these groups. The gap emerges only in the post-COVID period, appearing across both traditional invigilated exams and the online-style out-of-term formats. This indicates that the re-emergence of in-person assessment may have coincided with, or interacted with, broader recovery-phase factors—such as level of attendance, engagement, uneven support structures, or differential re-adjustment to on-campus academic expectations—rather than reflecting assessment format in isolation. In other words, the return to face-to-face assessment occurred under conditions very different from the pre-pandemic baseline, making it difficult to attribute the post-COVID gap to modality alone.

Control variables demonstrate remarkable stability across model specifications (Table 26), with gender, disability, socioeconomic status, and module-level effects showing minimal variation between in-term coursework, out-of-term examinations, and pooled analyses. This consistency increases confidence that the post-COVID ethnic divergence—visible across all assessment contexts and concentrated specifically among Black students—is not solely the result of model specification or covariate imbalance. The return to traditional assessment conditions is likely to interact with broader recovery-phase factors—such as attendance patterns, engagement, support, and differences in students' ability to re-establish effective learning routines—rather than reflect assessment format in isolation.

Table 26: Additional Model Effects - Control Variables and Interactions, all Models (Economics)

Categories	Overall	In-term	Out of term
A. ASSESSMENT & MODULE EFFECTS			
Out-of-term timing (main effect)	**0.251*** **		
× Pre-COVID	**-0.421*** **		**-1.882*** **
× Post-COVID	**-0.365** **		**-3.660*** and -3.331*** **
Core module	**1.894*** **	**1.423*** **	**1.907*** **
Term 2 vs Term 1 assessment	**-0.622*** **	**-0.563** **	**-0.592** **
Module level 2	**4.038*** **	**3.677*** **	**3.151*** **
Module level 3	**3.680*** **	**3.105*** **	**3.4178*** **
B. GENDER EFFECTS			
Female (main effect)	**1.835***	**1.632***	**1.619***
Female × Quantitative	**1.631***	**1.642** **	**1.601** **
C. DISABILITY CONDITION EFFECTS			
Any disability (COVID baseline)	**4.278***	**4.204** **	**4.387*** **
× Pre-COVID	**-9.012** **	**-9.160*** **	**-10.316*** **
× Post-COVID	-4.547	-4.282	**-4.308* and -6.864** **
D. SOCIOECONOMIC EFFECTS			
High professional occupation	**2.978** **	**2.980*** **	**2.515** **
Constant	63.986	64.655	64.166
E. RANDOM EFFECTS (SD)			
Student intercept	8.055	7.850	7.620
Core module slope	6.487	6.156	3.649
Module level slopes	7.486; 8.915	8.374; 6.558	7.537; 5.104
Residual	7.345	7.449	8.135

Note: *** p<0.001, ** p<0.01, * p<0.05. Reference: COVID period, male, no disability, non-professional background, in-term, non-quantitative, non-core modules, module level 1.

5.5 Conclusions of QMUL (Economics)

The evidence shows that attainment gaps largely disappeared during the COVID phase. Black and White students performed at similar levels, and the only noticeable difference was a temporary performance peak for Asian students, particularly in quantitative modules. The disadvantage–type gap therefore closed during COVID. It re-emerged only afterwards, when campuses reopened and emergency support was withdrawn, and only for Black students.

These gaps are not associated with any specific assessment mode but are present regardless of whether the task is quantitative, qualitative, written, oral, coursework, or examination.
Module domain, however, matters for understanding why some gaps do not open. Asian students show a clear quantitative advantage relative to their own non-quantitative performance. White and Black students do not. As a result, quantitative coursework prevents a White–Asian attainment gap from opening. But it does not explain the White–Black divergence, because Black students' decline is visible in both quantitative and non-quantitative modules.

From a policy perspective, this points towards a different set of priorities than those typically emphasised in assessment-design debates. The evidence suggests that the risk point is not

the crisis mode itself, but the transition out of crisis — when temporary supports are removed, expectations revert to normal, and students must self-regulate again. That is when Black students fall behind. This implies that interventions should prioritise support and scaffolding during recovery periods (e.g. return to campus, return to timed exams, and return to standard workload expectations), rather than seeking to standardise or redesign assessment formats in general. Early warning systems, proactive check-ins, and structured academic support during these transition phases are likely to be more effective than broad format reform.

These conclusions are subject to several constraints: Confounding variables present a challenge to causal interpretation. While the regression models control for module level, subject domain, gender, disability, and socioeconomic status, many potentially important factors remain unmeasured. Changes in student composition over time through admissions, variation in course selection patterns, differences in living arrangements during and after COVID, and differential access to private tutoring or family support could all contribute to the observed patterns. The analysis cannot definitively isolate COVID-related effects from other concurrent trends or policy changes.

Additionally, assessment comparability across periods remains uncertain: if examination difficulty, grading standards, or format changed in ways not fully captured by the format variable, this could influence the observed trajectories. Many institutions adopted more lenient grading during COVID, potentially masking learning gaps that only became visible when standards were restored.

Nevertheless, the analysis has clear advantages. It exploits multiple assessment contexts within the same institution (coursework, exams, quantitative and non-quantitative modules; pre-COVID, COVID, and post-COVID periods), applies consistent coding rules across formats, and holds institutional assessment policy constant. This provides an unusually direct way to observe whether gaps vary by assessment mode, delivery format, or task type. The fact that results converge across all of these contexts strengthens confidence in the core interpretation: the post-COVID widening is not an artefact of measurement or specific assessment choices but reflects a genuine divergence in outcomes concentrated on Black students after the Covid crisis.

5.6 All Courses Models (Selected Programmes)

We analyse now a larger sample that includes single honour degree in Economics and those joint degrees[13] that are comparable with the ones included in the Sussex's Department analysis. The added courses are Economics, Finance and Management, and Economics and Finance with their respective Placement versions.

Although Economics is included within the overall dataset and it represents 68% of the student sample used in the models, this comparison remains meaningful because the purpose is not to contrast fully independent samples, but to assess whether the patterns observed in the single Economics degree are distinctive relative to the broader assessment environment. Beyond differences in optional module choices, Economics students may experience a different balance of quantitative modules, a different sequencing of intermediate theory courses, and fewer externally-taught or interdisciplinary modules than students enrolled in different programmes.

[13] The only joint programme of the sample is Economics, Finance and Management which is run between the School of Economics and Finance and the School of Business and Management. The programme of Economics and Finance remains within the School of Economics and Finance.

As Economics forms part of the full sample, the differences reported here are likely conservative; excluding Economics would make the contrasts with other joint degrees even more pronounced.

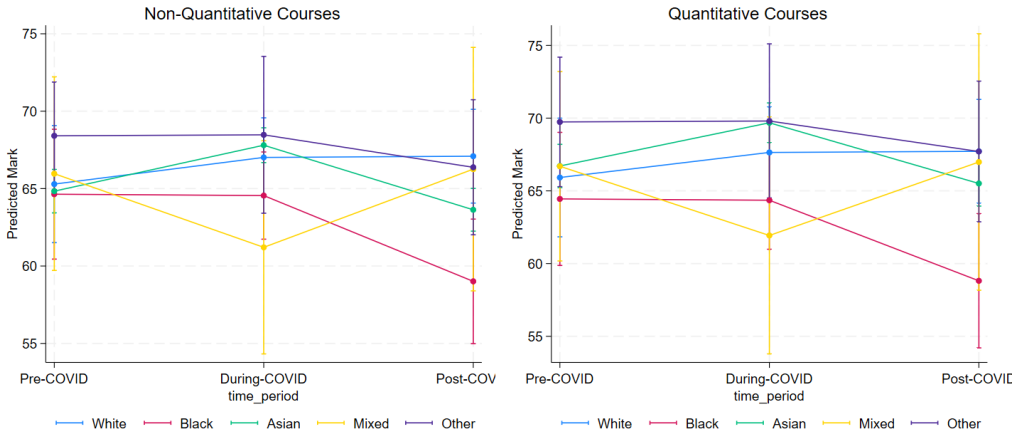
5.7 All Assessment Types: Overall Model (Selected Programmes)

The figures 16 and 17 show predicted marks for different ethnic groups across pre-COVID, during-COVID, and post-COVID periods for non-quantitative and quantitative modules. While some groups move closer together during COVID, this convergence is not universal. Mixed-heritage students experience a pronounced decline during COVID in both panels, performing well below all other groups, indicating that the emergency digital environment did not benefit all groups equally. White, Asian, and Other students show relatively stable or slightly improved performance during COVID, while Black students remain broadly similar to their pre-COVID levels. The convergence therefore mainly reflects the behaviour of White, Asian, and Other groups, with Mixed students underperforming and Black students showing little movement.

In the post-COVID period, the clearest pattern is the sharp decline in Black student performance across both non-quantitative and quantitative courses, creating a wide gap relative to White, Asian, and Other groups. Mixed-heritage students recover partially after COVID but do not fully return to their pre-COVID outcomes in non-quantitative modules. The similarity of the trajectories across the two panels suggests that these divergences are not driven by quantitative content but reflect broader shifts in assessment conditions and the return to traditional, higher-stakes formats.

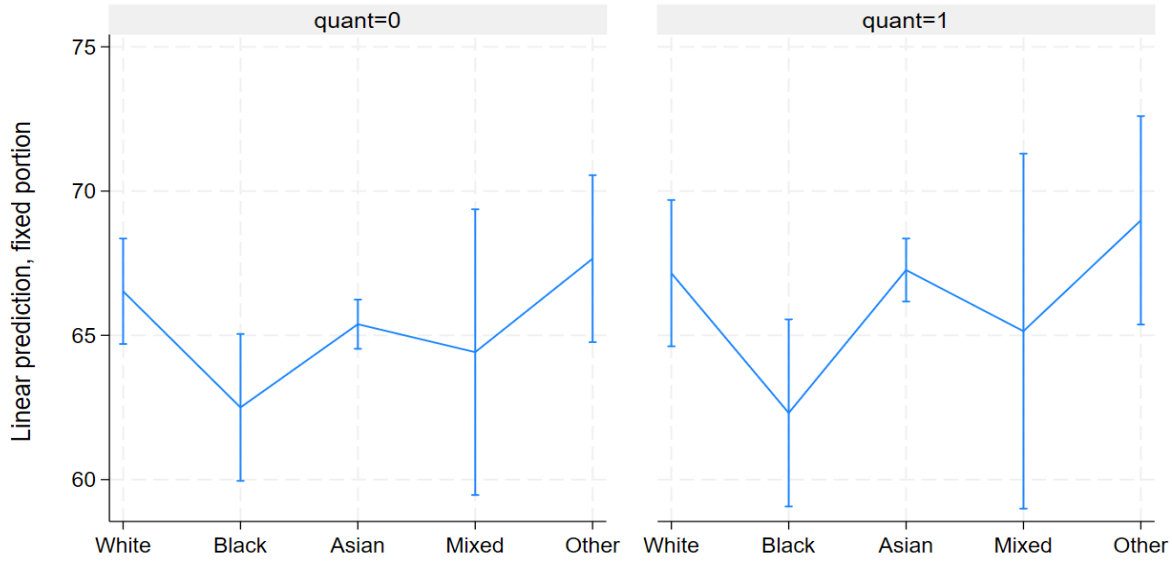
Compared with the Economics-only analysis, several parallels and differences emerge. The post-COVID decline for Black students appears in both datasets, confirming that this is a structural pattern rather than a discipline-specific effect. The Mixed-heritage group behaves differently in the wider School: they show a marked decline during COVID, whereas in Economics-only they do not exhibit this sharp dip. Asian students display a much clearer quantitative advantage in Economics-only than in the wider School, which helps explain why White–Asian gaps remain small or negative in Economics quantitative assessments. In both datasets, however, post-COVID widening of gaps occurs across coursework and exam types, reinforcing the conclusion that the divergence is driven by broader structural conditions rather than differences in subject matter.

Figure 16: Ethnic Attainment Trajectories Across COVID Phases (Selected Programmes)



Note: Vertical bars show 95% confidence intervals.

Figure 17: Ethnic Attainment gaps in Non-quantitative and Quantitative modules (Selected Programmes)



Note: Vertical bars show 95% confidence intervals.

Interestingly, Asian students do not display a comparative advantage in Quantitative modules within the non-Economics degrees—unlike what we observed earlier in the Economics single honour degree analysis (Table 21). This suggests that the stronger performance of Asian students in quantitative assessments is discipline-specific rather than a general pattern across the School.

Table 27: Ethnicity gaps by Non-quantitative and Quantitative Modules (Selected Programmes)

	White-Black Gap	p-value	White-Asian Gap	p-value	White-Mixed Gap	p-value	White –Other Gap	p-value
NON-QUANTITATIVE MODULES Gaps								
Pre	0.65	0.822	0.46	0.825	-0.67	0.857	-3.12	0.235
COVID	2.47	0.202	-0.79	0.572	5.81	0.123	-1.46	0.614
Post	**8.08*** **	0.002	**3.46** **	0.038	0.83	0.847	0.71	0.793
QUANTITATIVE MODULES Gaps								
Pre	1.47	0.639	-0.79	0.722	0.14	0.971	-3.82	0.216
COVID	3.29	0.161	-2.04	0.237	5.71	0.202	-2.16	0.492
Post	**8.9*** **	0.003	2.21	0.26	0.74	0.88	0.01	0.997
(*) not all joints degrees are included (we included only those related to Finance and Management).								

The comparison between Economics-only and other degree programmes reveals several noteworthy differences in post-COVID awarding patterns.

First, the White–Black gap is noticeably smaller in the other degrees than in the Economics-only sample, indicating that the larger gap observed at School level is driven predominantly by the Economics discipline itself.

Second, unlike the Economics-only analysis, the other programmes show a clear post-COVID White–Asian gap, suggesting that Asian students do not retain their earlier relative advantage outside Economics and may be more affected by the assessment formats or learning environments prevalent in other disciplines.

Finally, the White–Mixed gap observed elsewhere disappears entirely in the non-Economics sample, pointing to a more equitable post-COVID landscape for Mixed-heritage students in these programmes.

Together, these patterns underscore that the dynamics of awarding gaps are highly discipline-specific: some inequalities are amplified within Economics, while others emerge or dissipate only when looking beyond the discipline.

Focusing on White–Black and White–Asian differences within quantitative and non-quantitative modules is particularly meaningful, as these two groups of students together represent around 90% of the sample. Two patterns stand out.

First, the White–Black gap appears in both quantitative and non-quantitative modules, with only small and statistically insignificant differences between them, suggesting that quantitative content does not systematically amplify or reduce this gap.

Second, Asian students do not show a relative advantage in quantitative modules outside the Economics single-honours degree. In the wider School sample, the White–Asian gap remains small, negative, and not statistically significant across both module types. This contrasts with the Economics-only analysis, where Asian students performed notably better in quantitative modules. Taken together, these results indicate that the comparative advantage of Asian students in quantitative subjects is highly specific to the Economics programme and does not generalise to the broader set of degrees.

Table 28: Ethnic Attainment Gaps on Quantitative and Non-Quantitative modules (Selected Programmes)

	Non-quant	Quant	Quant Effect
White vs Black	2.47 (p=0.202)	3.29 (p=0.161)	0.82 (p=0.517)
White vs Asian	-0.79 (p=0.572)	-2.04 (p=0.237)	-1.25 (p=0.139)
Note: Positive = White advantage; Negative = Asian advantage. Quant Effect = additional gap from quantitative content. P-values in parenthesis.			

A simple comparison of gaps across quantitative and non-quantitative modules (the equivalent of Table 23 in the Economics-only analysis) shows that the relative advantage Asian students display in quantitative assessments appears to be specific to the Economics single-honours degree. In the wider School sample, the White–Asian gap remains small, negative, and statistically insignificant in both quantitative and non-quantitative modules, and the “Quant Effect” is close to zero.

Figure 18: Ethnic Attainment by In-term Assessments (Selected Programmes)

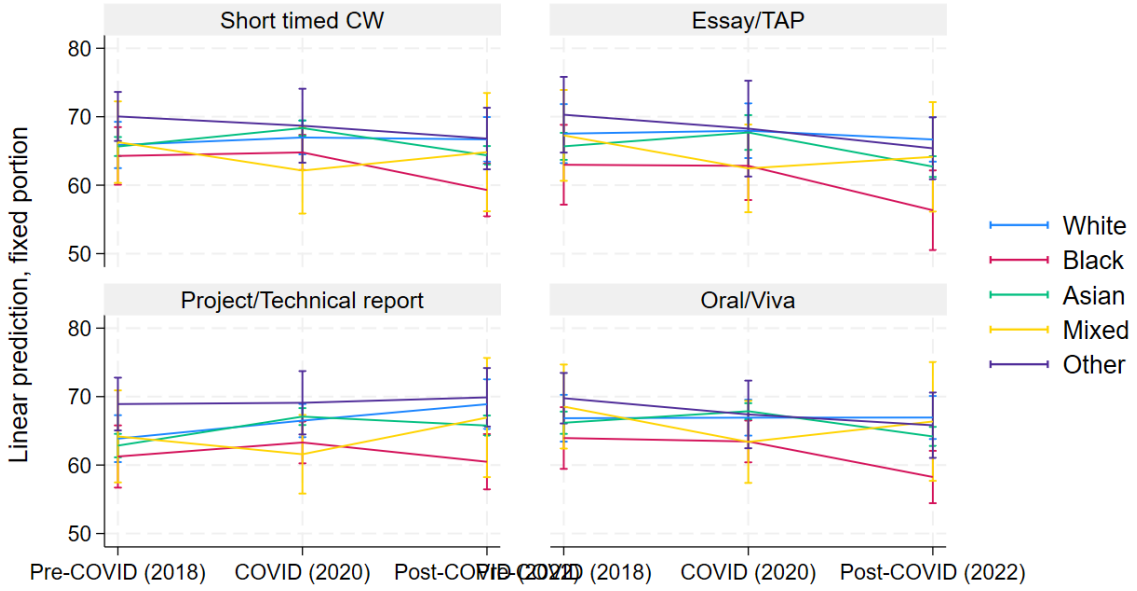


Figure 19: Ethnic Attainment Gaps Across COVID Phases by Assessment Type (Selected Programmes)

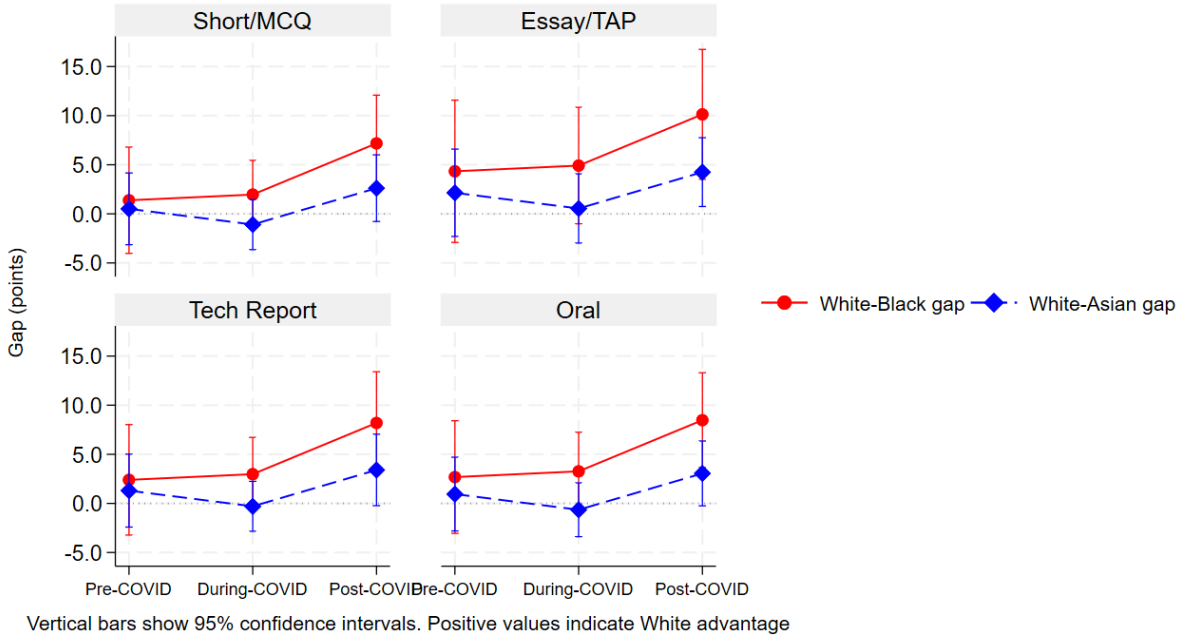


Table 29: Ethnic Attainment Gap, Coursework (Selected Programmes)

Assessment Type	Pre-Covid	p-value	COVID	p-value	Post-Covid	p-value
NON-QUANTITATIVE MODULES						
WHITE vs BLACK						
Short Timed CW	3.22	0.323	0.84	0.682	**9.26*** **	0.002
Essays/TAP	6.01	0.158	3.63	0.29	**12.05*** **	0.003
Projects/Tech	5.00	0.135	2.63	0.243	**11.04*** **	0.001
Oral/Viva	4.47	0.200	2.09	0.389	**10.51*** **	0.001
WHITE vs ASIAN						
Short Timed CW	1.93	0.396	-2.5	0.158	2.32	0.261
Essays/TAP	2.88	0.273	-1.54	0.478	3.27	0.123
Projects/Tech	3.58	0.11	-0.84	0.617	**3.97***	0.060
Oral/Viva	2.37	0.322	-2.06	0.273	2.76	0.17
QUANTITATIVE MODULES						
WHITE vs BLACK						
Short Timed CW	3.34	0.339	0.96	0.716	**9.37*** **	0.010
Essays/TAP	6.13	0.153	3.75	0.302	**12.16*** **	0.005
Projects/Tech	5.12	0.151	2.74	0.325	**11.16*** **	0.003
Oral/Viva	4.59	0.198	2.21	0.423	**10.62*** **	0.002
WHITE vs ASIAN						
Short Timed CW	-0.32	0.895	**-4.75** **	0.034	0.07	0.978
Essays/TAP	0.63	0.816	-3.79	0.13	1.02	0.679
Projects/Tech	1.33	0.575	-3.09	0.145	1.72	0.489
Oral/Viva	0.12	0.962	**-4.31***	0.059	0.51	0.832

A comparison of coursework awarding gaps across assessment types in quantitative and non-quantitative modules shows a markedly different pattern from the Economics-only results of Table 24. Before COVID, neither the White–Black nor the White–Asian gap is statistically significant in any assessment type, regardless of whether modules are quantitative or not. This contrasts with Economics single honours, where small but consistent gaps appear pre-COVID in several assessment formats.

During COVID, Black students show no coursework gaps in any module type, mirroring the equalisation seen in Economics-only; however, Asian students display clear advantages in quantitative assessments, a pattern that is absent in the non-quantitative modules. This advantage for Asian students in quantitative coursework is not observed in the wider School sample outside Economics, confirming that the strong quantitative performance of Asian students in the Economics-only degree is discipline-specific.

The post-COVID period shows the sharpest divergence between Economics and the other programmes. Black students experience large and highly significant awarding gaps across all assessment types in both quantitative and non-quantitative modules, indicating that the post-COVID widening is universal and not format-specific. This pattern is similar to Economics-

only, but the gaps here are even larger—suggesting that joint and non-Economics degrees amplify the White–Black gap relative to Economics single honours.

For Asian students, the picture differs sharply from Economics. Post-COVID White–Asian gaps emerge only in non-quantitative modules (notably in projects/technical tasks), while no gaps appear in quantitative coursework, where Asian performance is statistically indistinguishable from White students. In effect, quantitative modules appear to shield Asian students from the emergence of post-COVID awarding gaps, which aligns with the patterns displayed in the quantitative coursework graph. This protection is not observed in the Economics-only analysis, where Asian students generally return to parity across most assessment types.

Taken together, these results show that Economics single honours tend to compress gaps, while joint and non-Economics degrees tend to amplify them. The discipline-specific nature of assessment structure, quantitative intensity, and pedagogical design therefore plays a central role in shaping the size and direction of awarding gaps after COVID.

Figure 20: Ethnic Attainment by In-term Assessments (Selected Programmes)

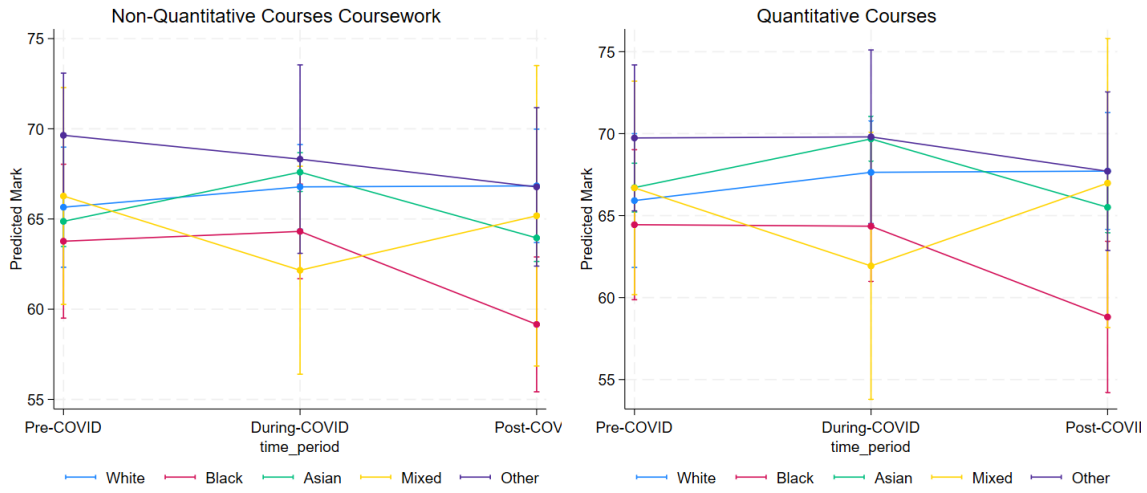


Figure 21: Ethnic Attainment by Type of Modules, In-term Assessments (Selected Programmes)

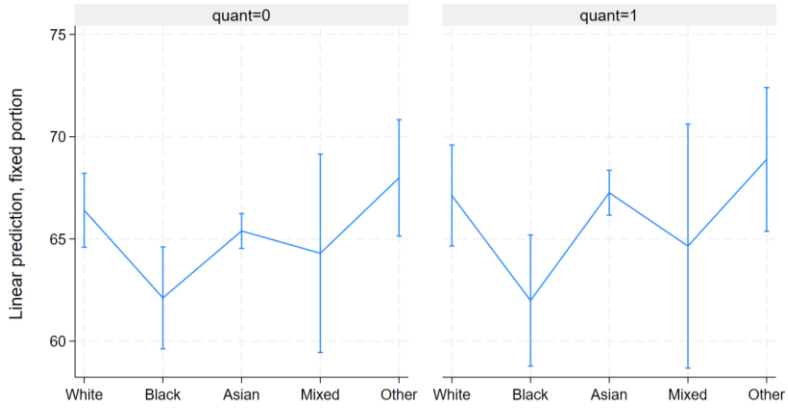


Figure 22: Ethnic Attainment, Final Exams (Selected Programmes)

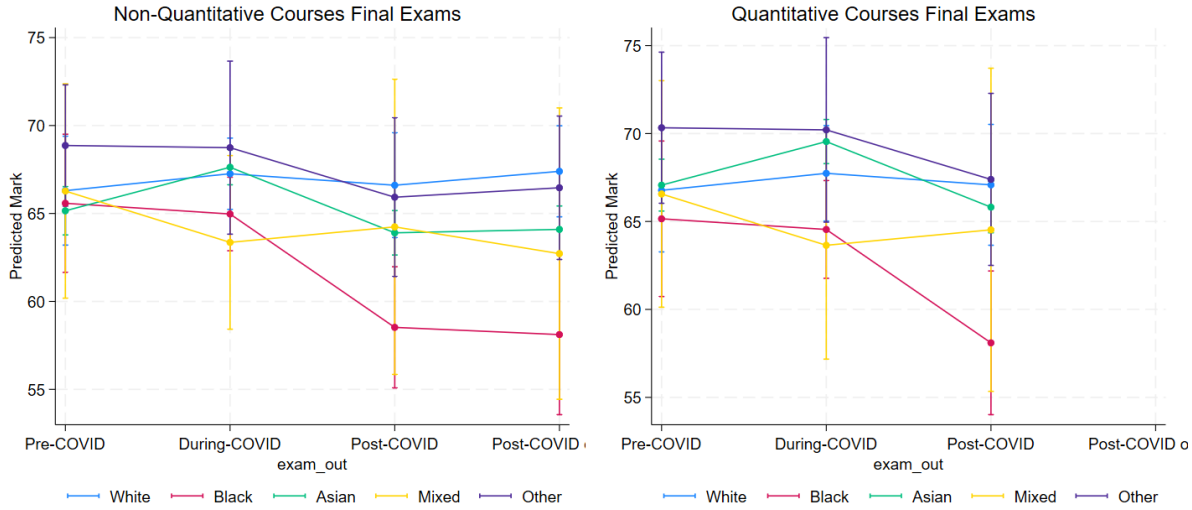
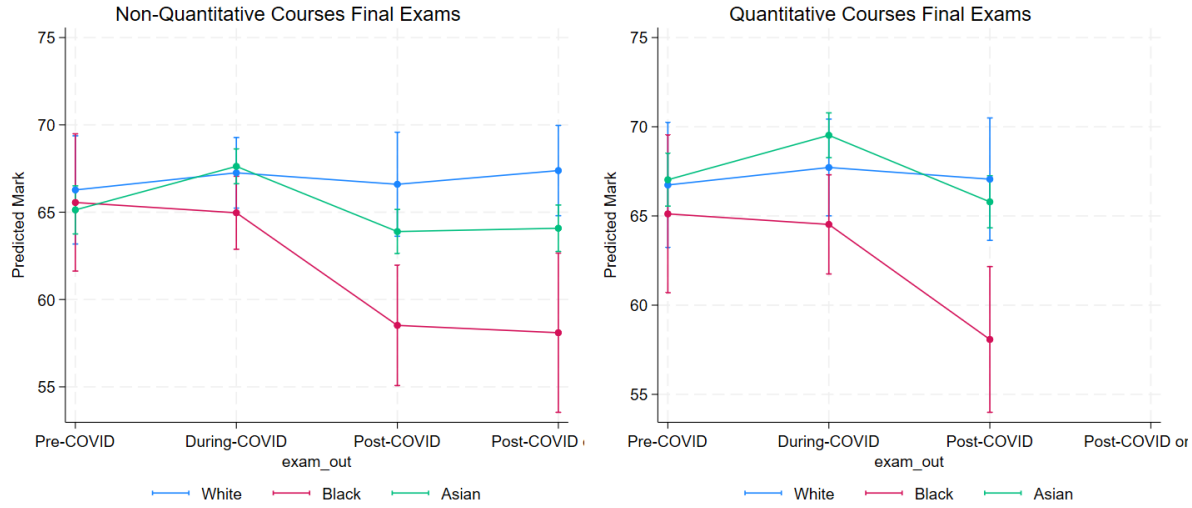


Figure 23: Ethnic Attainment in Final Exams (Selected Programmes)



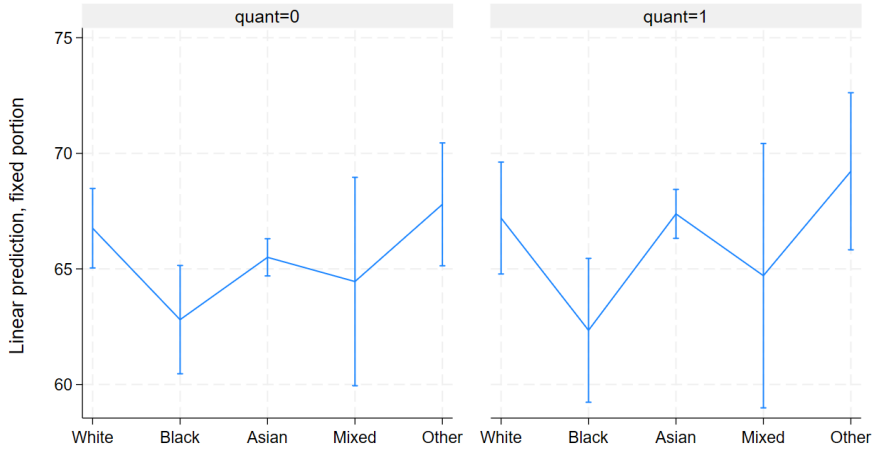
Compared with the Economics-only results, this graph shows clearer and more stable confidence intervals for Black and Asian students, making their post-COVID patterns easier to interpret. The tighter intervals confirm the sharp post-COVID decline for Black students and the absence of a White–Asian gap in quantitative finals. In contrast, Mixed and Other groups continue to show wide intervals, reflecting smaller sample sizes and greater uncertainty.

Table 30: Ethnic Attainment Gap, Final Exams (Selected Programmes)

Exam Type	White-Black	p-value	White-Asian	p-value	White-Mixed	p-value	White-Others	p-value
NON-QUANTITATIVE MODULES								
Traditional UEX pre-Covid	0.72	0.779	1.14	0.509	0.02	0.995	-2.57	0.276
COVID Digital	2.29	0.125	-0.37	0.747	3.9	0.155	-1.48	0.585
UEX Post-Covid	**8.08*** **	0.001	**2.71***	0.097	2.37	0.605	0.68	0.806
Untimed Post-Covid	**9.28*** **	0.001	**3.3****	0.024	4.68	0.293	0.94	0.705
QUANTITATIVE MODULES								
Exam Type	White-Black	p-value	White-Asian	p-value	White-Mixed	p-value	White-Others	p-value
Traditional UEX Pre-Covid	1.62	0.572	-0.29	0.88	0.21	0.956	-3.56	0.208
COVID Digital	3.19	0.109	-1.81	0.235	4.09	0.255	-2.47	0.412
UEX Post-Covid	**8.98*** **	0.001	1.27	0.502	2.56	0.612	-0.31	0.919

Compared with the Economics-only analysis, the final-exam results for single and joint degrees show a much clearer pattern of post-COVID widening. In both quantitative and non-quantitative modules, there are no significant gaps pre-COVID or during COVID, consistent with Economics-only. However, post-COVID the gaps in the wider School are larger and more widespread. In non-quantitative modules, sizeable White–Black and White–Asian gaps emerge in both UEX and untimed digital exams, whereas in Economics-only these gaps were either smaller or absent. In quantitative modules, the reintroduction of UEX assessments produces a strong post-COVID White–Black gap, again larger than in Economics-only, while no meaningful White–Asian gap appears. Overall, the joint and non-Economics degrees amplify the post-COVID awarding gaps: post-COVID gaps arise for Black and Asian students in non-quantitative modules, and for Black students in quantitative UEX exams, confirming your observation that the widening is more pronounced outside Economics.

Figure 24: Predictive Margins of Ethnicity Gaps, Final Exams (Selected Programmes)



The controlling factors in the selected programme sample largely replicate the patterns observed in the Economics-only analysis. Assessment timing, core status, module level, and the female advantage all behave in a similar way, as does the stronger performance of female

students in quantitative modules. The main point of divergence concerns disability: in Economics, disability showed a strong negative effect pre-COVID and a clear positive shift during COVID, whereas in the School-level results the pre-COVID penalty remains sizeable, but the COVID-period improvement is weaker and less consistent. Post-COVID disability effects are also more muted than in Economics-only. Overall, the broader sample displays the same structure of effects, but with a smaller and less volatile disability impact.

The coefficients for the Econ, Finance and Management (EFM) degree show that students enrolled in this programme systematically achieve lower marks than single-honours Economics students, even after controlling for assessment type, module level, timing, gender, disability, and socioeconomic background. This suggests that the EFM programme has structural features—such as a broader curriculum mix, different assessment cultures, and varying levels of quantitative preparation—that may place students at a relative disadvantage compared with those in single-honours Economics.

Table 31: Overall, In-Term and Out-of-term models (Selected Programmes)

Categories	Overall	In-term	Out-of-term
A. ASSESSMENT & MODULE EFFECTS			
Out-of-term timing (main effect)	**0.275*** **		
× Pre-COVID	**-0.444*** **		**-1.882*** **
× Post-COVID	**-0.223** **		**-3.66*** and -3.331*** **
Core module	**1.593*** **	**1.4172*** **	**1.626*** **
Term 2 vs Term 1 assessment	**-0.791*** **	**-0.750*** **	**-0.645*** **
Module level 5 (year 2)	**3.782*** **	**3.677*** **	**2.687*** **
Module level 6 (year 3)	**3.298*** **	**3.105*** **	**3.008*** **
Course (single honour ref .cat)			
Econ, Finance and Management	**-3.826** **	**-4.127*** **	**-3.623** **
Econ and Finance	-0.446	-0.591	-0.585
B. GENDER EFFECTS			
Female (main effect)	**2.305*** **	**2.127*** **	**2.159*** **
Female × Quantitative	**1.553** **	**1.629*** **	**1.800*** **
C. DISABILITY CONDITION EFFECTS			
Any disability (COVID baseline)	1.336	1.103	1.92
× Pre-COVID	**-6.243** *	**-6.296** *	**-8.138** **
× Post-COVID	-1.094	-0.667	-0.822 and -4.612
D. SOCIOECONOMIC EFFECTS			
High professional occupation	**2.468*** **	**2.522*** **	**2.052** **
Constant	64.281	65.032	64.625

Note: *** p<0.001, ** p<0.01, * p<0.05. Reference: COVID period, male, no disability, non-professional background, in-term, non-quantitative, non-core modules, module level 1.

5.8 Conclusions of QMUL (Selected Programmes)

The QM results show that ethnic attainment gaps shift substantially across periods and assessment types, with notable differences compared to the Economics-only analysis. Three key conclusions emerge:

1. **Pre-COVID and COVID periods show minimal awarding gaps.** Across both quantitative and non-quantitative modules, attainment gaps between ethnic groups remain small and statistically insignificant before and during COVID. This is consistent with Economics-only results and reflects the temporary equalisation associated with emergency online assessment conditions.
2. **Post-COVID gaps widen sharply, particularly in final examinations.** After the return to stable assessment formats, large White–Black gaps emerge across both quantitative and non-quantitative final exams, and White–Asian gaps appear in non-quantitative finals. These post-COVID disparities are larger in the wider sample than in Economics-only, indicating that joint and non-Economics programmes contribute most to the widening pattern.
3. **Coursework reveals discipline-specific differences not visible in Economics-only.**
   The comparative quantitative coursework advantage observed for Asian students in Economics does not appear in the broader QM sample, where quantitative coursework does not systematically favour or disadvantage any group. Non-quantitative coursework, however, displays post-COVID awarding gaps, echoing the pattern seen in final examinations.

Across structural and background factors, the wider sample broadly replicates the patterns observed in Economics-only, with a few notable differences. Female students continue to outperform male students, including in quantitative modules, and performance rises predictably from Levels 4 to 6 in both datasets. Core modules are again associated with higher marks, and students from high professional backgrounds achieve consistently stronger results. Assessment timing also behaves similarly: out-of-term and digital conditions confer small advantages during COVID but not before or after. The main divergence concerns disability. In Economics-only, disability shows a strong pre-COVID penalty and a clear positive shift during COVID, suggesting that emergency online conditions were particularly beneficial. In the wider School, the pre-COVID penalty remains sizeable, but the COVID-period improvement is weaker and less consistent, and post-COVID effects are more muted. Overall, structural patterns are stable across the institution, but disability effects display more volatility within Economics than in the Selected Programmes broader sample.

5.9 Limitations of QMUL analysis (Selected Programmes)

- **Small subgroup sizes**
  Some ethnic groups—particularly Mixed-heritage and “Other”—have small sample sizes, leading to wider confidence intervals and less precise estimates. This issue is more pronounced in Economics-only, where overall cohort sizes are smaller.
- **Differences in programme structure**
  Non-Economics BSc degrees included in the analysis sit within the same School but follow different module combinations and assessment formats. Joint programmes involve modules delivered by different Schools across the university. These structural differences may influence attainment patterns and limit direct comparability with Economics-only results.

- **Variation in COVID-period assessment adaptations**
  Economics relied more consistently on quantitative and standardised digital formats, while the wider sample used a broader mix of coursework and exam types. Variation in emergency online provision may contribute to divergent COVID-period effects.
- **Aggregated disability categories**
  Disability is grouped under "any disability," masking differences across specific conditions. Declaration rates also vary, which may partly explain why disability effects differ between Economics-only and the Selected Programmes sample.
- **Unobserved student characteristics**
  Administrative data do not include information on study habits, digital access during COVID, external responsibilities, health issues, or engagement. These unmeasured factors may influence performance differently across programmes.
- **Selection and progression patterns**
  Economics-only students' progress through a more standardised curriculum, whereas students across the wider School follow diverse pathways, potentially introducing selection effects not captured in the models.
- **Period grouping**
  Grouping assessments into pre-COVID, COVID, and post-COVID periods smooths over year-to-year variation and may mask cohort-specific dynamics.

5.10 Policy Implications for QMUL (Selected Programmes)

The results for Economics and the wider Selected Programmes sample point to several areas where assessment practices and student support could be strengthened. Three broad implications emerge:

1. **Strengthen resilience and equity in assessment design.**
   The post-COVID widening of gaps—especially in high-stakes final exams—indicates that some formats systematically disadvantage particular groups. Disciplines should review whether shorter, structured, time-limited formats (which were associated with reduced gaps during COVID) can be adopted more widely, and whether out-of-term assessments and concentrated exam periods can be better balanced to reduce pressure points.
2. **Address variation across programmes and Schools.**
   Differences between Economics-only, non-Economics BSc programmes, and joint degrees suggest that variation in assessment culture plays a role in attainment patterns. Greater alignment of expectations across quantitative modules, clearer coordination between Schools contributing to joint programmes, and calibration of marking practices would help reduce structural inconsistencies.
3. **Target academic support at key groups and transition points.**
   Persistent post-COVID White–Black gaps in examinations point to the need for enhanced assessment literacy, structured revision opportunities, and clearer performance expectations. Level 5 remains the most challenging progression point across programmes, indicating a need for strengthened academic skills support. Socioeconomic disparities also remain visible, suggesting the value of expanded mentoring, study skills provision, and targeted support for students from lower-SES backgrounds.

6. Chapter 4: Case Study in the University College London IMB Programme

Introduction

Attainment gaps by ethnicity persist in UK higher education and may be sensitive to how learning is assessed. The pandemic precipitated rapid changes to assessment conditions (for example, remote or open-book examinations), creating an opportunity to examine attainment patterns across different assessment modes (coursework, examinations, mixed of coursework and examinations) under differing operational contexts. In this case study, we investigate whether the BAME-White gap varies by assessment mode and period, and how each group's typical performance relates to the 60% classification benchmark.

Methods

The analysis draws on 10,496 anonymised module-level outcome data from the UCL IMB programme, covering 615 students across 19 taught modules between 2016 and 2022, comprising 3,485 records for White students and 7,011 for BAME students. For comparability, observations are grouped into three periods: pre-COVID (up to and including 2018), during-COVID (2019 to 2021), and post-COVID (2022). Each module is classified by its primary assessment mode as coursework, examination, or mixed (coursework plus examination). Within each ethnicity, period, and mode combination, we summarise performance using the 60th percentile of marks. The attainment gap is defined as White minus BAME (percentage points). To anchor practical significance, we compare each summary with the 60% classification boundary.

Results

Pre-COVID distributions (see Figure 25, Panels a and b) show that typical performance for both groups lies above the 60% benchmark across all modes. In coursework and mixed assessments, the margins above 60% are broadly comparable, whereas examinations display a higher typical mark for BAME students than for White students. Specifically, the 60th percentile in BAME examinations exceeds the White value, yielding a negative gap for this period.

During-COVID distributions (see Figure 25, Panels c and d) indicate an overall compression of differences between groups. The distance between group 60th percentiles narrow in examinations, approaching parity, while mixed assessments remain tightly clustered across ethnicities. Coursework shows a modest White advantage in this period, although typical marks for both groups remain above 60%.

Post-COVID distributions (see Figure 25, Panels e and f) reveal a reversal in examinations: the typical White examination mark rises relative to the BAME mark, creating a positive gap. Coursework margins above 60% soften for both groups, and mixed assessments continue to show small differences, with both groups' typical marks located comfortably above the benchmark.

Figure 25: Distribution of Student's Marks by Assessments Mode and COVID phases

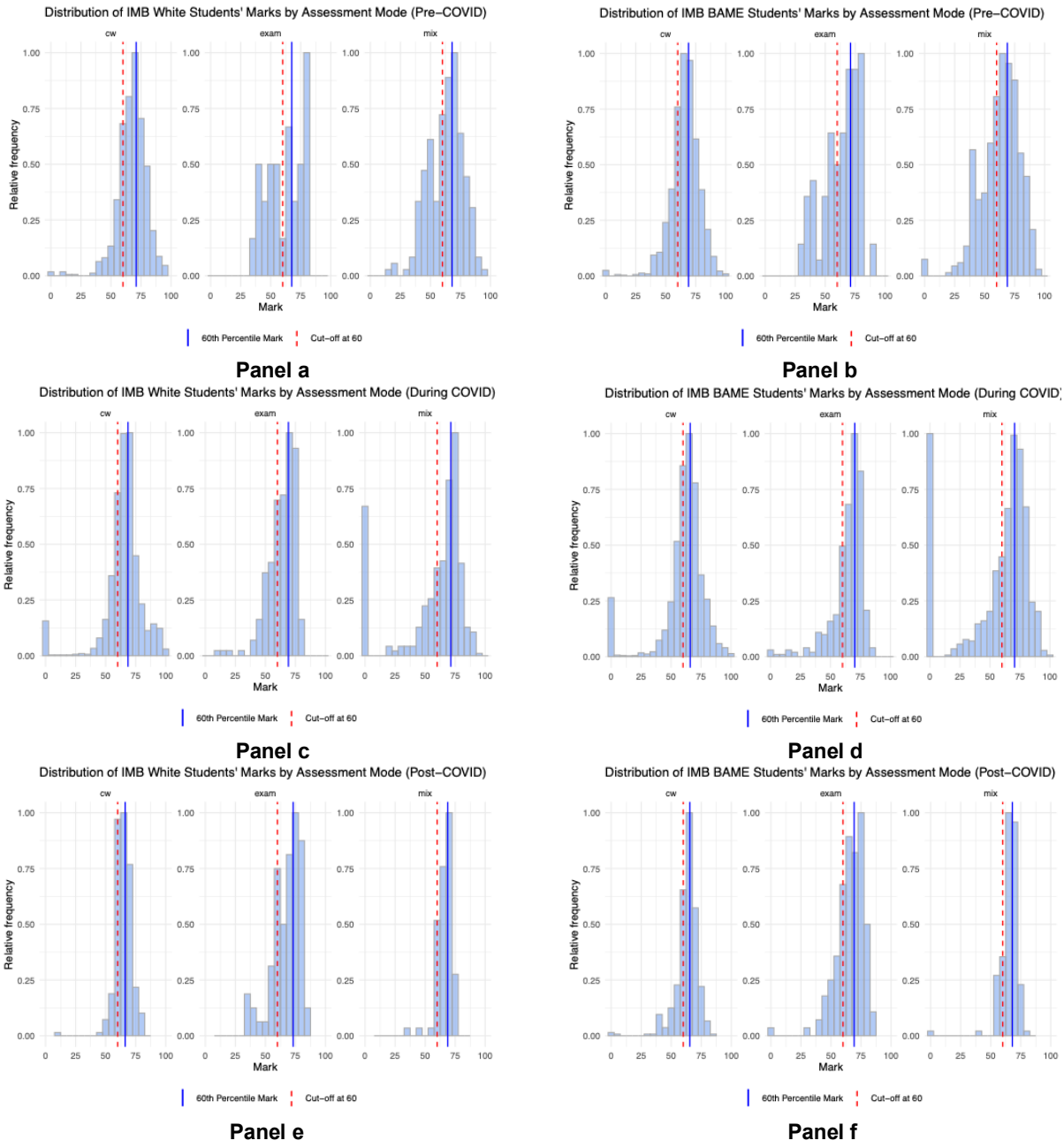


Table 32 reports the attainment gaps numerically. The examination mode exhibits the largest temporal swing: a BAME advantage pre-COVID (-3.6 percentage points) moves towards near parity during COVID (-1.0), and becomes a White advantage post-COVID (+3.8), implying a net shift of 7.4 points across the observation window. Coursework gaps are comparatively small and narrow in the post-COVID period, while mixed-mode gaps remain close to zero throughout.

Table 32: Attainment Gaps (White minus BAME, in percentage points) by Period and Assessment Mode

Period	Coursework	Examination	Mixed
Pre-COVID	+2.0	-3.6	-0.9
During-COVID	+2.6	-1.0	+0.9
Post-COVID	+0.8	+3.8	+0.8

Beyond the gaps, the relationship between typical performance and the 60% benchmark is informative. Pre-COVID, typical marks exceeded 60% in all cases. In examinations, BAME students had the largest margin (approximately eleven percentage points above 60%), whereas the White examination 60th percentile sat closer to the threshold. During COVID, the margin above 60% narrowed for coursework in both groups, while remaining comparatively high in examinations and mixed assessments. Post-COVID, coursework margins softened further for both groups; examination performance diverged, with the White 60th percentile rising to more than thirteen points above 60% and the BAME 60th percentile remaining around nine points above 60%. Mixed assessments remained stable for both groups, several points above the benchmark.

Discussion

The evidence indicates that assessment conditions are consequential for equity. The reversal observed in examinations from a BAME advantage pre-COVID to a White advantage post-COVID is consistent with changes in exam delivery and constraints over the period, such as the transition from remote or open-book formats during the pandemic to more traditional invigilated settings thereafter. Mixed assessments are associated with persistently small gaps, which may reflect the benefits of triangulating evidence of attainment and reducing dependence on a single performance context. Coursework gaps were modest and, by the post-COVID period, had nearly closed. While coursework can, in principle, amplify disparities in time availability and academic capital (e.g., familiarity with exemplars and effective use of office hours), the post-COVID near-closure suggests that contemporaneous practices, such as clearer rubrics and digital feedback workflows, enhanced during COVID may have mitigated these effects.

Implications

The findings suggest three practical directions for assessment design and programme monitoring. First, stabilising examination parameters across years, including time limits, item formats, and permitted resources, may reduce volatility in gaps as delivery modes change. Second, retaining well-designed mixed assessments appears promising for equity while maintaining strong performance above the 60% benchmark. Third, the post-COVID transition from a pre-COVID BAME advantage in examinations to a White advantage warrants targeted investigation. Plausible contributors include the return to invigilated, time-constrained conditions that heighten sensitivity to speeded recall and high-stakes stress, uneven recovery from pandemic-related disruptions that disproportionately affects single-sitting exams and shifts in cohort composition.

7. Chapter 5: Final Conclusions of the Study

Across the three institutions considered—QMUL, UoS and UCL—the analysis shows that ethnic attainment gaps are not fixed characteristics of student groups but shift in response to assessment design, learning conditions, and wider structural changes associated with the COVID-19 period. Despite differences in programme structures, cohort composition and assessment cultures, several robust cross-institution patterns emerge.

1. **COVID-19 temporarily reduced inequality, particularly through digital timed assessments.**
   At all three universities, ethnic attainment gaps narrowed or disappeared during the COVID year. Black–White gaps closed, Asian students performed strongly, and the smallest gaps occurred in structured, time-limited digital examinations. This suggests that the emergency online environment—reduced stakes, simplified formats and stronger scaffolding—lowered barriers for groups that typically face structural disadvantage.
2. **The return to standard assessment practices produced renewed widening of gaps.**
   In both institutions, the post-COVID period saw a sharp re-emergence of ethnic gaps as assessment formats stabilised, and pre-pandemic expectations resumed. The most consistent pattern is the post-COVID divergence between Black and White students, which appears across coursework and final exams, and across quantitative and non-quantitative modules. This indicates that the renewed gap is not tied to a single assessment mode but reflects broader pressures associated with returning to autonomous, high-stakes assessment environments.
3. **Traditional in-person exams and extended open-ended tasks generate the largest disparities.**
   Across QMUL, UoS and UCL, the post-COVID widening of ethnic gaps is most evident in traditional invigilated final exams (UEX) as well as in extended assessments such as essays, written reports and untimed digital finals. These formats share a reliance on sustained independent study, high cognitive load and strong exam-taking skills, all of which appear to disadvantage certain groups—particularly Black students—more sharply in the post-COVID period. By contrast, structured, time-limited digital assessments during COVID generated much smaller gaps.
4. **Quantitative performance patterns do not consistently protect Asian students from post-COVID divergence.**
   While Asian students performed strongly during COVID across institutions, their post-COVID outcomes do not show a uniform quantitative advantage. In QMUL Economics-only, some quantitative coursework advantages appear, but these do not generalise to UEX or to the wider Selected Programmes sample. At QMUL and UoS, Asian students' post-COVID performance often aligns with or falls slightly below White students, and White–Asian gaps re-emerge in several assessment types. Quantitative assessment therefore does not explain the post-COVID widening between White and Black students, nor does it consistently shield Asian students from post-COVID declines.
5. **Post-COVID performance fell across all ethnic groups.**
   QMUL, UoS and UCL all show broad declines in predicted marks after COVID, reflecting wider structural challenges such as disrupted schooling, reduced study routines, cost-of-living pressures and increased academic expectations. Although the magnitude varies—Black and Mixed-heritage students typically declining most sharply—the direction of change is consistent.
6. **Intersectional patterns reveal persistent structural vulnerabilities.**
   Across universities, students with mental-health-related disabilities experienced

sustained and significant disadvantages. Socioeconomic effects were also visible: students from more privileged backgrounds tended to achieve higher marks, though the strength of this pattern varied. Gender patterns were similar in direction across institutions: female students consistently outperformed male students. These intersectional findings reinforce that ethnicity is only one dimension of structural inequality within assessment.

7. **The key vulnerability lies in the transition out of crisis, not the crisis itself.** A shared insight is that gaps did not widen during the COVID emergency but after it, when emergency support was withdrawn and high-stakes assessment practices resumed. This post-crisis divergence—especially affecting Black students—appears linked to challenges in self-regulation, independent study, time management and navigating intensified expectations. This suggests that recovery phases require as much institutional attention as crisis periods.

In summary, the findings demonstrate that attainment gaps are structurally sensitive and highly responsive to changes in assessment design and learning conditions. Both institutions show that inequalities narrow when assessments are structured, time-limited, and well scaffolded, and widen when students must rely on independent study and extended high-stakes tasks. The post-COVID environment has therefore become a renewed point of vulnerability for structurally disadvantaged groups, especially Black students, highlighting the need for sustained, evidence-based reforms to assessment practices, academic support structures, and transitional scaffolding.

8. Common Policy Implications

1. Assessment format is a major driver of inequality across both institutions: structured, time-limited tasks are associated with smaller gaps, while extended, open-ended assessments consistently generate the largest disparities, particularly for Black students.
2. The post-COVID widening of gaps suggests that the return to traditional assessment formats coincided with a broader structural change in the learning environment, one that interacts with socioeconomic background, disability, and other forms of structural disadvantage; strengthening scaffolding, assessment literacy, and academic support is therefore essential.
3. Improving coherence across programmes and Schools—especially in joint degrees—together with enhanced, targeted support for students facing structural barriers will be necessary to build a more equitable and resilient assessment landscape.

9. Future Research

Several areas of future research would meaningfully extend the insights from this study and help institutions respond to an evolving assessment landscape.

First, there is an urgent need for more granular institutional data, particularly on ethnicity and background characteristics. Current reporting categories, especially aggregated “BAME” labels, are too broad and mask important differences within Asian, Black, Mixed-heritage, and other groups. More detailed ethnicity data, combined with richer information on socioeconomic background, disability type, schooling history, and prior attainment, would allow for more precise and actionable analysis.

Second, the study would benefit from better module-level information on assessment design and teaching practices. With the current dataset, it is not possible to differentiate the cognitive demands, scaffolding, or feedback structures of specific assessments, nor to capture

variations across Schools, especially in joint degrees. Strengthening institutional datasets on assessment characteristics would make it possible to identify which design features most consistently widen or narrow attainment gaps.

Third, although Generative AI was not used during the period examined in this report, its rapid expansion in subsequent years introduces a new structural factor for future research. As a growing share of assessments are completed off-campus and without invigilation, institutions will need to understand how AI availability interacts with assessment format, whether uptake differs across student groups, and whether AI risks amplifying or mitigating existing inequalities.

Finally, the strong post-COVID widening of gaps highlights the need to examine the broader structural shifts in students' lives that may have emerged after the pandemic. Changes in commuting patterns, increased work hours, cost-of-living pressures, caring responsibilities, and reduced access to quiet study space may all affect engagement with high-stakes assessments. Longitudinal and qualitative research into these evolving living and study conditions, alongside academic confidence, study habits, and perceptions of assessment fairness, would help explain why some groups were disproportionately affected during the recovery period and inform more effective support and assessment design.

Together, these areas of inquiry would provide institutions with a clearer and more comprehensive evidence base, supporting assessment and policy decisions that remain fair, inclusive, and robust in the face of ongoing technological and structural change.

10. Bibliography and Suggested References

The following references were consulted during the preparation of this report and may offer further insights into the topic.

Advance HE. (2020). *Ethnicity awarding gaps in UK higher education in 2019/20*.

Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment and Evaluation in Higher Education*, *39*(2), 205–222. https://doi.org/10.1080/02602938.2013.819566

Bloxham, Sue., & Boyd, Pete. (2007). *Developing effective assessment in higher Education: a practical guide*. Open University Press.

Bottan, D., McKee, D., Orlov, G., & McDougall, A. (2022). Racial and gender achievement gaps in an economics classroom. *International Review of Economics Education*, *40*. https://doi.org/10.1016/j.iree.2022.100239

Brown, S., & Sambell, K. (2023). Rethinking assessment by creating more authentic, learning-oriented tasks to generate student engagement. In *Formative and Shared Assessment to Promote Global University Learning* (pp. 150–167). IGI Global. https://doi.org/10.4018/978-1-6684-3537-3.ch008

Bunce, L., King, N., Saran, S., & Talib, N. (2021). Experiences of black and minority ethnic (BME) students in higher education: applying self-determination theory to understand the BME attainment gap. *Studies in Higher Education*, *46*(3), 534–547. https://doi.org/10.1080/03075079.2019.1643305

Cagliesi, G., & Ghanei, M. (2022). Team-based learning in economics: Promoting group collaboration, diversity and inclusion. In *Journal of Economic Education* (Vol. 53, Issue 1, pp. 11–30). Routledge. https://doi.org/10.1080/00220485.2021.2004276

Cagliesi, M. G., Hawkes, D., & Smith, S. (2023). Narrowing awarding gaps: the contributory role of policy and assessment type. *Studies in Higher Education*, *48*(11), 1665–1677. https://doi.org/10.1080/03075079.2023.2209597

Cotton, D. R. E., Joyner, M., George, R., & Cotton, P. A. (2016). Understanding the gender and ethnicity attainment gap in UK higher education. *Innovations in Education and Teaching International*, *53*(5), 475–486. https://doi.org/10.1080/14703297.2015.1013145

Eldakhakhny, B., & Elsamanoudy, A. Z. (2023). Discrimination Power of Short Essay Questions Versus Multiple Choice Questions as an Assessment Tool in Clinical Biochemistry. *Cureus*. https://doi.org/10.7759/cureus.35427

Elsamanoudy, A., Shehata, M., Almarabheh, A. and Alrefaie, Z. (2024). Evaluation of modified essay questions (MEQs) as an assessment tool in third-year medical students' modular summative assessment. BMC Medical Education, 24(1), p.1445. https://doi.org/10.1186/s12909-024-06469-w

Engelhardt, B., Johnson, M., & Meder, M. E. (2021). Learning in the time of Covid-19: Some preliminary findings. *International Review of Economics Education*, *37*. https://doi.org/10.1016/j.iree.2021.100215

Fawns, T., Bearman, M., Dawson, P., Nieminen, J. H., Ashford-Rowe, K., Willey, K., Jensen, L. X., Damşa, C., & Press, N. (2025). Authentic assessment: from panacea to criticality. *Assessment and Evaluation in Higher Education*, *50*(3), 396–408. https://doi.org/10.1080/02602938.2024.2404634

Islam, K., Ahmadi, P., & Yousaf, S. (2017). *Assessment Formats and Student Learning Performance: What is the Relation?* Available at*: https://arxiv.org/abs/1709.07948*

Mahmud, A., & Gagnon, J. (2020). Racial disparities in student outcomes in British higher education: examining Mindsets and bias. *Teaching in Higher Education*, *28*(2), 254–269. https://doi.org/10.1080/13562517.2020.1796619

Mountford-Zimdars, A., & Moore, J. (2017). *Addressing Awarding Gaps through Assessment Design*.

Paredes Fuentes, S., Burnett, T., Cagliesi, G., Chaudhury, P., & Hawkes, D. (2023). *Who Studies Economics? An Analysis of Diversity in the UK Economics Pipeline*.
Popoola O., Tavasci D., & Ventimiglia, L. (2022). *An attainment and awarding gap analysis for ethnic groups/gender and assessment type in the School of Economics and Finance QMUL*. Report for SP110b Assessment & Feedback Workstream.
Richardson, J. (2008). *Degree attainment, ethnicity and gender: a literature review*.
Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. In *Higher Education* (Vol. 35), pp. 453-472. https://doi.org/10.1023/A:1003196224280.
Singh, G. (2009). *Synthesis A synthesis of research evidence. Black and minority ethnic (BME) students' participation in higher education: improving retention and success*. www.heacademy.ac.uk/evidencenet
Thomas, L., & May, H. (2010). *Inclusive learning and teaching in higher education*.
Walker, R. (2021). *Awarding gaps and assessment: a briefing paper for EAC*.

11. Appendices

Appendix A.1: Descriptive Statistics, UoS (Economics)

Variables	Pre-COVID 2018.19	COVID 2020.21	Post-COVID 2022.23	Total (Pooled)
Number of Students	228	220	207	655
Ethnicity				
White	69%	64%	65%	66%
Black	11%	9%	8%	9%
Asian	10%	15%	13%	12%
Mixed	8%	10%	10%	9%
Other	2%	3%	4%	3%
Total	100%	100%	100%	100%
Gender				
Male	74%	77%	77%	76%
Female	26%	23%	23%	24%
Total	100%	100%	100%	100%
Socioeconomic				
No Free School Meal	94%	90%	92%	92%
Free School Meal	6%	10%	8%	8%
Total	100%	100%	100%	100%
Declared Disabilities				
No Disability	81%	85%	78%	81%
Cognitive Disability	4%	5%	6%	5%
Mental health	9%	5%	7%	7%
Multiple Disabilities	3%	2%	3%	3%
Physical	2%	4%	4%	3%
Social Disability	1%	0%	2%	1%
Total	100%	100%	100%	100%
Modules				
Non-quant	76%	72%	77%	74%
Quant	24%	30%	23%	26%
Total	100%	100%	100%	100%

Placement				
Non-placement	91%	92%	85%	90%
Placement	9%	8%	15%	11%
Total	100%	100%	100%	100%
Progression level				
Level 4	40%	31%	30%	34%
Level 5	38%	35%	32%	35%
Level 6	22%	34%	38%	31%
Average assessment per student (*)	16	18	16	16
CWS	8	10	9	9
Final exams	8	8	7	7
All assessments	4,679	5,982	4,473	15,134
In -Term Assessments (sub sample)	2,321	3,387	2,413	8,121
Short Time Assessments	25%	31%	41%	32%
Essay	29%	20%	27%	25%
Written Reports	6%	10%	12%	10%
Technical Tasks	15%	10%	7%	10%
Interactive and Engagement Task	25%	29%	13%	23%
Final Exams (sub sample) (**)	2,312	2,594	2,050	6,956
UEX Traditional	79%	-	-	
(Digital) Timed	5%	34%	50%	
(Digital) Untimed	16%	66%	50%	

(*) computed including the overlapping of students; (**) Technical Reports and Group Work were excluded from the final-exam categories because they accounted for only 47 cases in total (0.6%) and could not be meaningfully grouped with the main final-exam formats.

Appendix A.2: Descriptive Statistics, UoS (3 Departments)

Variables	Pre-COVID	COVID	Post-COVID	Total
	2018.19	2020.21	2022.23	Pooled
Number of Students	617	662	832	2,111
Departments				
Management	44%	49%	51%	48%
Accounting and Finance	20%	18%	24%	21%
Economics	37%	33%	25%	31%
Ethnicity				
White	71%	66%	65%	67%
Black	8%	7%	8%	8%
Asian	11%	15%	15%	14%
Mixed	8%	9%	8%	8%
Other	2%	3%	4%	3%
Total	100%	100%	100%	100%
Gender				
Male	68%	69%	72%	70%
Female	32%	31%	28%	30%
Total	100%	100%	100%	100%
Socioeconomic				
No Free School Meal	92%	92%	91%	91%
Free School Meal	8%	8%	9%	9%
Total	100%	100%	100%	100%
Declared Disabilities				
No Disability	83%	83%	81%	82%
Cognitive Disability	6%	8%	8%	7%
Mental health	6%	3%	4%	5%
Multiple Disabilities	3%	3%	3%	3%
Physical	2%	3%	3%	3%
Social Disability	0%	0%	1%	0%
Total	100%	100%	100%	100%

Modules				
Non-quant	90%	88%	91%	90%
Quant	10%	12%	9%	10%
Total	100%	100%	100%	100%
Placement				
Non-placement	80%	84%	82%	82%
Placement	20%	16%	18%	18%
Total	100%	100%	100%	100%
Progression level				
Level 4(*)	39%	36%	32%	35%
Level 5 (*)	39%	36%	37%	37%
Level 6 (*)	22%	28%	31%	28%
Average assessment per student (*)	15	17	16	16
CWS	8	10	10	10
Final exams	7	7	6	6
All assessment	11,924	16,996	17,355	46, 275
In -Term Assessments (sub sample)	5,966	9,690	10,545	26,201
Short Time Assessments	25%	28%	37%	31%
Essay	28%	20%	20%	22%
Written Reports	14%	19%	19%	18%
Technical Tasks	9%	9%	5%	7%
Interactive and Engagement Task	25%	24%	19%	22%
Final Exams (sub sample) (**)	5,847	7,305	6,805	19,957
UEX Traditional	71%	-	10%	
(Digital) Timed	9%	32%	35%	
(Digital) Untimed	20%	68%	55%	

(*) computed including the overlapping students; (**)Technical Reports and Group Work were excluded from the final-exam categories because they accounted for a very small percentage and could not be meaningfully grouped with the main final-exam formats.

Appendix A.3: Results of UoS (3 Departments)

Variables	Overall sample		Coursework		Final Exam	
A. ACADEMIC FACTORS	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Out of term (Final exams) vs Coursework	-4.061***	0.000	-	-	-	-
Out of Term (High stake FE) vs Coursework	-6.617***	0.000	-	-	-	-
Progression Level 5 vs Level 4	-0.439	0.315	0.810	0.110	3.468***	0.000
Progression Level 6 vs Level 4	0.986**	0.012	4.667***	0.000	4.392***	0.000
Quantitative Module vs non Quant.	0.086	0.836	2.463***	0.000	-2.199***	0.000
Placement vs non-Placement	4.367***	0.000	4.101***	0.000	3.857***	0.000
Time: COVID (2020)	-0.181	0.732	1.140			
Time: Post-COVID (2022)	-4.019***	0.000	-8.554***	0.000	-	-
Accounting and Finance vs Management	-0.276	0.637	-1.885***	0.002	-0.987	0.133
Economics vs Management	0.486	0.349	-1.901***	0.001	0.843	0.117
B. SOCIOECONOMIC						
FSM vs non-FSM (Pre-COVID)	-5.753***	0.003	-4.023**	0.026	-7.722***	0.001
FSM × Time: COVID	5.146**	0.024	2.543	0.232	7.086***	0.008
FSM × Time: Post-COVID	4.618*	0.061	2.470	0.280	5.819**	0.047
C. SEX: Female vs Male	3.072***	0.000	3.358***	0.000	2.444***	0.000
D. DISABILITY EFFECTS						
Disability: Cognitive vs no Disability	-0.985	0.185	-0.645	0.431	-1.272	0.171
Disability: Mental Health	-5.363***	0.000	-4.970***	0.000	-4.889***	0.000
Disability: Multiple	-2.485	0.092	-2.052	0.117	-1.725	0.315
Disability: Physical	0.527	0.553	0.887	0.354	0.1789	0.859
Disability: Social	-0.976	0.866	-0.066	0.993	-2.251	0.598
E. RANDOM EFFECTS (SD)						
Student intercept	5.559		5.995		7.172	
Module slope (quant)	6.499		5.125		4.848	
Module level slopes (Level 4)	8.504		9.414		7.281	
Module level slopes (Level 5)	7.377		7.218		9.556	
Time period slopes (COVID)	6.388		5.204			
Time period slopes (Post-COVID)	8.531		10.331			
Out of term assessment slope	3.859					
Residual	17.726		17.906		15.011	

Note: * p<0.10, ** p<0.05, *** p<0.01

Appendix B.1: Descriptive Statistics, QMUL (Economics)

Variables	Pre-COVID (2018.19)	COVID (2020.21)	Post-COVID (2022.23)	Total (Pooled)
Number of Students	174	132	191	497
Ethnicity				
White	15%	11%	16%	15%
Black	14%	9%	10%	11%
Asian	63%	73%	62%	65%
Mixed	4%	5%	5%	4%
Other	4%	2%	7%	5%
Total	100%	100%	100%	100%
Gender				
Male	63%	69%	71%	68%
Female	37%	31%	29%	32%
Total	100%	100%	100%	100%
Socioeconomic				
All other Occupations	86%	93%	84%	87%
Highest Occupation	14%	7%	16%	13%
Total	100%	100%	100%	100%
Declared Disabilities				
No Disability	91%	96%	95%	94%
Any Disability	9%	5%	5%	6%
Total	100%	100%	100%	100%
Modules				
Non-quant	70%	77%	79%	75%
Quant	30%	24%	22%	25%
Total	100%	100%	100%	100%
Placement				

Non-placement	100%	99%	93%	97%
Placement	0%	1%	7%	3%
Total	100%	100%	100%	100%
Progression level				
Level 4	32%	37%	38%	36%
Level 5	41%	34%	35%	37%
Level 6	27%	28%	27%	28%
Total	100%	100%	100%	100%
Average assessment per student (*)	18	20	21	21
CWS	57%	64%	63%	62%
Final exams	43%	36%	37%	38%
All assessments	3,622	4,160	4,907	12,689
Terms(**)				
Autumn	54%	57%	58%	57%
Spring	46%	43%	42%	43%
Total	100%	100%	100%	100%
In-Term Assessments (sub sample)	2,061	2,679	3,105	7,845
Short Time Assessments	80%	60%	61%	66%
Essay/TAP	2%	1%	17%	8%
Project/Technical Reports	9%	24%	10%	14%
Oral/Viva	10%	15%	12%	12%
Total	100%	100%	100%	100%
Average number of CWK assessments per student	10	13	13	13
Final Exams (sub sample)	1,553	1,480	1,802	4,835
UEX Traditional	100%	0%	0%	
Digital (COVID)	0%	100%	0%	

Digital Timed (Post)	0%	0%	77%	
Digital Untimed (Post)	0%	0%	23%	
Total	100%	100%	100%	100%
Average number of Final Exams per student	8	7	8	8

(*) computed including the overlapping of students. (**)Term 1 and Term 2 (together) modules were excluded from the sample used in the regressions because they accounted for only 32 cases in total (0.25%) and could not be meaningfully grouped with the main final-exam formats.

Appendix B.2: Descriptive Statistics, QMUL (Selected Programmes)

Variables	Pre-COVID (2018.19)	COVID (2020.21)	Post-COVID (2022.23)	Total (Pooled)
Number of Students	245	213	278	736
				Ethnicity
White	16%	14%	18%	16%
Black	12%	9%	9%	10%
Asian	63%	72%	62%	65%
Mixed	6%	2%	4%	4%
Other	4%	4%	6%	5%
Total	100%	100%	100%	100%
Gender				
Male	65%	67%	75%	69%
Female	36%	33%	25%	31%
Total	100%	100%	100%	100%
Socioeconomic				
All other Occupations	84%	91%	83%	86%
Highest Occupation	16%	9%	17%	15%
Total	100%	100%	100%	100%
Declared Disabilities				

No Disability	93%	95%	95%	94%
Any Disability	7%	5%	5%	6%
Total	100%	100%	100%	100%
Modules				
Non-quant	73%	79%	78%	77%
Quant	27%	21%	22%	24%
Total	100%	100%	100%	100%
Placement				
Non-placement	98%	96%	90%	94%
Placement	2%	4%	10%	6%
Total	100%	100%	100%	100%
Progression level				
Level 4	33%	34%	35%	34%
Level 5	42%	37%	38%	39%
Level 6	25%	29%	26%	27%
Total	100%	100%	100%	100%
Average assessment per student (*)	17	20	20	20
CWS	57%	64%	63%	62%
Final exams	43%	36%	37%	38%
All assessments	5,135	6,161	6,645	17,941
Terms (**)				
Autumn	54%	56%	58%	56%
Spring	46%	44%	42%	44%
Total	100%	100%	100%	100%
In-Term Assessments (sub sample)	2,946	3,921	4,183	11,050
Short Time Assessments	79%	62%	63%	67%

Essay/TAP	2%	1%	16%	7%
Project/Technical Reports	10%	21%	9%	14%
Oral/Viva	10%	16%	13%	13%
Total	100%	100%	100%	100%
Average per student	10	13	13	10
Final Exams (sub sample) (**)	2,189	2,240	2,462	6,891
UEX Traditional	100%	0%	0%	32%
Digital (COVID)	0%	100%	0%	33%
Digital Timed (Post)	0%	0%	78%	28%
Digital Untimed (Post)	0%	0%	22%	8%
Total	100%	100%	100%	100%
Average per student	7	7	7	7

(*) computed including the overlapping of students. (**)Term 1 and Term 2 (together) modules were excluded from the sample used in the regressions because they accounted for only 32 cases in total (0.18%) and could not be meaningfully grouped with the main final-exam formats.

Gabriella Cagliesi
Business School, University of Sussex

Deyu Ming
School of Management, University College London

Susan Smith
School of Management, University College London

Valeria Terrones
Business School, University of Sussex

Luigi Ventimiglia
School of Economics and Finance, *Queen Mary University of London*

This report was developed through a cross-institutional collaboration led by the University of Sussex, with key contributions from Queen Mary University of London and University College London (UCL) **and funded by the QAA through its Collaborative Enhancement Projects scheme.**