How we Assess Mathematics Degrees: Post Covid-19 Era

Ana Rivera Montero
Maria Chistodoulidou Manitara
Prapti Maitra

Year 4 Project
School of Mathematics
University of Edinburgh
March 3, 2025

Abstract

Two previous studies mapped the assessment methods used in universities across the UK, ranging throughout the league table positions. These indicated a strong reliance on closed book examinations, with a slight decrease on the follow-up study, possibly due to the growth of alternative modules within mathematics. The study described in this report replicated the earlier work, to identify any changes in the assessment diet, particularly due to the Covid-19 pandemic. The main finding from the analysis is the strong reliance on examinations for assessment, although now with the additional appearance of open book examinations, likely influenced by the pandemic's impact on assessment practices. The report also addresses a small case study focusing on the University of Edinburgh which revealed a statistically significant difference in the examination methods pre and post Covid in contrast to the whole UK sample.

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(*Ana Rivera Montero*
Maria Chistodoulidou Manitara
Prapti Maitra)

Contents

Chapter 1

Introduction

C. K. Y. Chan (2023) describes that assessment serves four main purposes, including, examining achievement, promoting student learning, overseeing accountability both for funding and for the public and assuring and maintaining quality of academic standards. With these qualities in mind, we can understand the importance of continually investigating the methods employed for assessment; as these will give an insight into the educational standards of those that provide it. This is specially important in the context of higher education, where students get prepared to be lifelong learners and gain competencies for their employability profile (C. K. Chan & Luk, 2022).

In recent years, academia has undergone a period of change, adaptation and innovation due to the recent worldwide pandemic, caused by Covid-19 (*See Section* 2.3.2). This global emergency, that forced everyone to remain on lockdown, had a particular effect on higher education. The UK, amongst other countries, saw measures across the community such as, suspended face-to-face teaching, encouraged working from home and most importantly in the context of this report, change in examination arrangements (Crawford et al., 2020). On the 5th of May 2023, The World Health Organization held a press conference to announce the end of the global emergency status (WHO, 2023). At this point, restrictions have been lifted and normal in-person assessment can be held. However, in the higher educational system, many of the changes that were put into place during the switch to online teaching, has now motivated educators to apply changes in the methods used (Trevisan et al., 2021).

As final year university students, we have lived through all of these shifts in previously established educational model, specifically within the department of mathematics. In particular, our university experiences have developed from having modules fully online to then a gradual adaptation back to in-person lectures and assessment. Taking our perspective of the last 4 years as a catalyst, our motivation for this study stems from an interest into a systematic analysis of how methods employed by universities now differ from previous times, which we have not experienced. As such, we have a close relationship with the study in hand, and find ourselves personally involved with the findings.

This study aims to provide a further analysis of the development of mathematical teaching in the field of assessment within higher education institutions in the UK. We will consider the work of Iannone and Simpson, 2011 and Iannone and Simpson, 2022, and provide a continuation of their research, using similar methods and approaches. The previous research examines the assessment methods used in the UK up until 2011, followed by a continuation in 2019; after a decade of changes in the higher education system, stemmed by the increased popularity of e-assessment (Iannone & Simpson, 2022). However, we now focus on replicating the work, in the context of 2024. This implies taking into account the effects of the online shift during the pandemic and discussing whether academia has reverted back to the model before Covid-19 or whether an adaptation has happened.

Given our close relationship with the University of Edinburgh as students, and the availability of resources, a case study comparing the assessment diet between 2019 and 2023 of this institution will be performed together with the general analysis.

Thus, with these ideas in mind, we revisit the research of the summative assessment and collect new data, within the context of a post-pandemic world. We will assess whether there have been any changes in the past 4 years, with a worldwide pandemic in between. We will address the same research questions as the previous studies, with the addition of a specified focus on post-pandemic changes and a special case study on the methods employed at the University of Edinburgh. The research questions are as follows:

1. What mix of assessment methods are observed in mathematics departments in the UK?

2. How are different assessment methods used in different topic areas?

3. Are there links between the assessment diet and the league table placements of university mathematics departments?

4. Are there any differences in the assessment diet at degree course level?

5. Have there been any changes to the mathematics assessment after the 2020 pandemic?

6. Are there any observed differences between the assessment diets in England and Scotland?

7. How does the focused study of one institution compare to the research on a sampled United Kingdom study?

Chapter 2

Literature Review

2.1 Introduction and Context

The literature on assessment at university level shows a clear need for a variety of assessment methods (e.g. Houston, 2001; Laursen and Rasmussen, 2019; Nortvedt and Buchholtz, 2018). This need is seen through many forms such as formalized policies, like The Subject Benchmark Statement for Mathematics, Statistics and Operational Research (MSOR) released by the Quality Assurance Agency in 2023; or by reports on experimenting in which ways students respond to assessment methods in order to better design undergraduate mathematics programs (e.g. Iannone and Simpson, 2015b; Shorter and Young, 2011; Voskoglou, 2019), amongst others. The Subject Benchmark Statement for MSOR states a need for good practice with assessment, relating to the availability of a variety of assessment methods. There is a special mention of the need to make reasonable adjustments to the assessment methods in response to the ”indivdual and the discipline, and provide authentic participation in all course activities” (QAA, 2023, p. 9). The community of mathematics also sees instances like the Mathematical Association of America (MAA) funding projects such as the three-year, “Supporting Assessment in Undergraduate Mathematics” (SAUM); where the main objective is to aid in the design of competent assessment programs in curricular blocks of undergraduate mathematics (Steen, 2006).

Smith and Wood (2000) pinpoint that this need for evaluating how mathematics is assessed at university stems from the assumption that “assessment drives what students learn” (p. 126). In this way, depending on the approach of assessment, pupils will take a more surface approach or in-depth approach to their understanding. Through this point of view, the kind of assessment technique applied by the teacher will directly tell students what is expected of them and what is valued in their learning. In particular, Houston (2001) explains that within mathematics, ”what you assess is what you get” (p. 408), meaning that generally, the nature of the teaching and learning will be controlled by the assessment tools.

It is found in university mathematics, that the main source of summative assessment consists of an unseen closed book examination (CBE) with timeconstraint (Houston, 2001). However, in recent years, there has been some discussion on whether this method of summative assessment should be the primary approach to use (e.g. Nortvedt and Buchholtz, 2018; Videnovic, 2017). In particular, some critics argue that the CBE, as it stands, does not test the abilities that some lecturers value and instead are "rewarding memory and diligence" (Iannone & Simpson, 2015a, p. 80); where other assessments such as project and oral examinations could potentially do. Moreover, this emphasis on tests and examinations is criticized for causing consequences on the curriculum taught and the teaching methods. Particularly, adapting the curriculum to teach easily tested material and increased focus on tests leading to teachers teaching for the sole purpose of preparing for an exam, as well as, raising doubts on the reliability of an exam as a test for student's abilities (Kulm, 2013). Further criticism on the traditional method stems from the idea that within a larger context, the work of a mathematician doesn't focus on working through time-constrained, unseen problems. The consolidation of a deeper learning of the subjects is ignored in favour of a lack of connection between topics and poor growth in mathematical curiosity (Houston, 2001).

Directions for future development in the field of mathematical assessment practices have been suggested, for example, "efforts to improve relation between research and practice" and "applying technology to develop better measures of mathematical competence" (Nortvedt & Buchholtz, 2018, p. 567).

Since the earlier reports of the transformation of assessment, the literature of the field gives an account of evolution of the use of technology in the classroom, reporting a change from its use for formative assessment (Panero & Aldon, 2016), towards an innovative use for summative assessment (Sangwin, 2019a). Following then, during the Covid-19 pandemic, all education was online, and with such, new technologies have been introduced into the classroom as a result of that (*See Section* 2.3.2).

2.2 Previous Studies

Given that we are replicating a study already performed twice before, we now take into consideration the previous reports on the assessment diet data by Iannone and Simpson (2011, 2022). Below we summarize the content of these two papers.

2.2.1 The summative assessment diet: how we assess in mathematics degrees

Iannone and Simpson (2011), providing data collected in 2010, address 1, 2, and 3 of the research questions presented in this report (*See Chapter* 1). It follows a systematic survey, similar to the one being performed in Chapter 3 in our study; for a total of 11 universities spread across the chosen league table in the UK. The findings highlighted the sustained dominance of closed-book examinations across the institutions and a strong correlation with the proportion of CBE contributing to the final degree and the university rankings. Furthermore, the report underlined the emergence of alternative assessment methods, particularly in modules emphasizing practical skills and historical perspectives (i.e. statistics, history of mathematics, mathematics education and final year projects).

The authors conclude that while there is a clear preference for closed book exams in mathematics departments, simultaneously alternative assessment methods are gaining ground, particularly in areas that require practical skills, historical perspectives, or an emphasis on employability. It was suggested that the heavy reliance on CBEs could be due to the simplicity of delivery, the demand of mathematical skills and the difficulty of academic misconduct (Iannone & Simpson, 2011).

2.2.2 How we assess mathematics degrees: the summative assessment diet a decade on

A decade later, the authors revisit the discussion of summative assessment in UK mathematics departments. In this context, Iannone and Simpson (2022) address that the literature in the field now focuses predominantly on the rising popularity of e-assessment and the use of dialogic forms of assessment. The research questions remain the same as the previous research.

Even though the report was published in 2022, the data in this study was collected in 2019 and thus, represent the state of the field before Covid. Again, the methods resemble those described in the methodology of this report; taking data from degrees coded 'G100' in the Higher Education Statistics Agency's coding system and systemically taking public information on the methods of assessment used by universities across the UK. Note that in this case, there were 23 universities included in the sampling with the exclusion of Scottish universities that didn't appear, without any particular intention. This report revealed a continued dominance of closed book exams, with only a marginal decrease. The authors suggest this decrease is due to the incorporation of additional modules, as a result of an increased focus on employability and skills modules.

Overall the comparison with the data from the previous study revealed that little has changed. There is full data set, in both occasions of the research, for 19 departments. These show a small decrease in the mean total percentage of closed book exams from 72% to 68% which, the report proposes, might be the result of the growth of alternative modules. However, the relationship of league department and proportion of closed book exams remained the same with the Russel Group universities dominating the Post-92 institutions in their use of closed book examinations. There was an increase in the presence of skills modules, programming and computational modules, as well as an increase on the weighting of grades attained by final year projects.

2.3 Background and Covid-19 Impact

We now present a chronological account of the state of the art in higher education mathematics assessment pre, during, and post Covid-19.

2.3.1 Pre Covid-19

For decades, the dominant and preferred method of assessment by mathematics educators has been the use of summative exams, typically closed-book timed exams. Closed-book exams are favoured for their ability to reduce cheating and their ease of creation and grading (Iannone & Simpson, 2011). However, this form of assessment lacks the valuable feedback provided by formative assessments to students.

Over the years, there has been significant criticism regarding the exclusive reliance on summative closed-book exams. As Houston (2001) discussed, learning mathematics primarily to pass examinations, often results in shallow understanding, focused solely on memorization rather than grasping the subject as a whole. The concept of time- constrained, unseen, written examination restricts the complexity of problems assessed in the exam, preventing opportunities for discussion, research, reflection, or the use of technology-essential skills for a mathematician's professional growth. A survey conducted by Williams and Wong (2009) reported that students expressed a preference for open-book, open-web (OBOW) exams over closed-book exams due to the flexibility in exam location and the opportunity for deeper learning. However, a significant concern regarding OBOW exams was the potential for academic misconduct, particularly cheating. Interestingly, the study indicated that the likelihood of cheating was comparable between OBOW exams and closed-book exams. In this way, the survey suggested that instead of focusing on preventing cheating by a minority, the priority should be on maximizing the benefits that the majority of students could get with OBOW exams.

In the years before the pandemic, when the concept of solely remote teaching was not widespread, researchers explored innovative methods of assessing mathematics beyond traditional closed-book exams. Investigations into the use of computer-aided assessment (CAA) to evaluate students' work were conducted (Sangwin, 2004). Sangwin (2019b) explored the possibility of transitioning paper-based linear algebra exams into electronic assessments. He achieved this by developing a linear algebra course examination for first year undergraduate students, confirming its viability. Additionally, he highlighted the benefits of electronic examinations, including improved reliability, efficiency in marking, and reduced costs. A growing popularity has been observed for e-assessment systems such as NUMBAS [1] and STACK [2] which are advanced Computer-Aided Assessment (CAA) systems designed for solving mathematics with emphasis on formative assessment. In addition to multiple choice questions, systems like STACK enable students to respond using

[1]https://www.numbas.org.uk/
[2]https://stack-assessment.org/

algebraic expressions to mathematical problems (Sangwin, 2015). These systems also offer the option of randomizing questions and provide immediate feedback to students. The response from students to this new technology varied. Initially, many encountered challenges, particularly with syntax barriers due to their unfamiliarity with notation conventions. However, most students quickly adapted and commented that they appreciate the immediate feedback available, which helps them learn from their mistakes.

In addition to the prevalent use of closed book exams, a variety of other assessment methods were also employed prior to the pandemic. Iannone and Simpson (2011) categorized these methods into eight groups: closed-book examination, dissertation, open-book examination, multiple-choice test, oral examination, regular example sheets, project and presentation. While all of these assessment methods were utilized, closed book exams typically constituted the largest proportion of the overall grade. This was emphasized in the follow up research, where, once again, Iannone and Simpson (2022) reported that CBEs remained the most used assessment method. However, it was revealed that this wasn't the case for all modules, including statistics, which also used open book exams for assessment. This could be attributed to the provision of a formula sheet to students, effectively categorizing these exams as open book. The reason behind this approach is likely the extensive list of formulas associated with the subject, allowing instructors to assess students' techniques rather than their memory recall (Iannone & Simpson, 2011, 2022).

Pre-Covid, traditional assessment methods in mathematics education predominantly relied on closed-book exams, supplemented by occasional experimentation with alternative approaches. As educators were gradually exploring innovative assessment techniques and integrating technology into mathematics teaching, they could not foresee the rapid shift that awaited them with the onset of the Covid-19 pandemic.

2.3.2 During Covid-19

During the pandemic and after the mandatory university closures, assessment methods had to adjust to accommodate the new circumstances. The transition to online teaching and assessment became necessary. Mathematics lecturers worldwide found themselves with limited time to adapt to the new situation, facing the added complexity of remotely teaching mathematics, due to the nature of the course, characterized by its reliance on symbolism and mathematical notation (Trenholm & Peschke, 2020). Hence we see issues arise when needing to explain mathematical concepts online without traditional classroom tools such as boards. Almost overnight teachers had to change their teaching style and adopt new pedagogies using unfamiliar tools and technology (Johann Engelbrecht, 2023). The use of slides proved to be ineffective in the teaching of mathematics, posing a challenge for educators as writing and illustrating mathematical concepts is integral to the teaching process of the course (Ní Fhloinn & Fitzmaurice., 2021b). To transition the symbolic and formal writings of mathematics to digital formats, most lecturers replaced the traditional blackboard and chalk that they were mostly using before the pandemic, to the use of a tablet. Others chose to prepare handwritten material prior to their lectures and present it during the class (Cassibba et al., 2020). Their adaptation to new technology occurred independently, with the majority demonstrating proficiency in basic digital skills such as presentations, yet a noticeable gap in formal training for online learning platforms. While comfortable with fundamental digital skills, lecturers faced challenges with more advanced skills like animation and video editing (Irfan et al., 2020).

Class delivery quickly incorporated a broader range of hardware, including webcams, external microphones, and stylus pens, alongside the traditional blackboards/whiteboards and laptops (Eabhnat Ní Fhloinn, 2021). Many mathematics lecturers chose to deliver the syllabus using a mixture of live online sessions and pre-recorded sessions, with the drawback of pre-recordings being the inability of students to ask questions live. The incorporation of live and recorded sessions during emergency remote teaching brought about a diversification of the software landscape, with communication systems like Zoom, Skype, MS Teams, and Blackboard Collaborate gaining wider usage.

There are mixed views on whether the teaching practices used during the pandemic were more beneficial or disadvantageous for students. Distant learning forced lecturers to prepare their material before the lecture to share during the class. Many lecturers found this approach significantly more efficient since they could use class time to explain better and with greater detail the concepts to students (Cassibba et al., 2020). Educators observed improved organization and planning in their lessons compared to their pre-pandemic practices. Also, since they prepared all the notes and published them to students after class, students didn't have to take notes of everything and could devote more attention on understanding the course material. On the contrary, most lecturers found it challenging to assess whether students were keeping up with the material, and felt a loss of interaction with their students.

A survey conducted during the pandemic to higher education institutions in UK and Ireland, revealed that all 36 of the respondents from their sample reported that all assessments were carried out remotely, with the majority of universities asking students to write and upload solutions to a traditional-style paper format (Henley & Hilliam, 2022). Approximately half of the respondents to the survey adopted a 'short release' assessment format, offering students comparable writing time as in an invigilated exam on campus, with an additional 20-60 minutes for electronic submission. Some departments used a '24-hour release' exam, allowing 24 hours for completion, or a '48-hour release' format. A small percentage opted for an 'intermediate release' style, giving students 6-9 hours to write and submit solutions. Lecturers preferred the short release to minimize the risk of academic misconduct, particularly as concerns arose that longer exams provided students with increased opportunities to search the internet for answers. In response, some professors adjusted their exam formats to prevent students from easily finding answers online.

Fhloinn (2021) reported that several educators observed notable variations in the overall grade distribution of students during the pandemic, compared to previous years. Teachers who found their students receiving higher marks claimed that this could be due to academic misconduct since the absence of in-person supervised closed-book assessment, increased the likelihood of cheating among students. Others claimed that assessments were also easier, perhaps as an adjustment to the challenges posed by remote learning and also students were given more time to complete the exam than they would have pre-pandemic. On the contrary, a proportion of teachers also found their students grades decreasing and claimed that this could be due to extra stress caused by the pandemic situation and increased difficulty of online exams (Fhloinn, 2021).

The perspective of mathematics students on remote teaching and assessment was examined via a survey conducted across Irish universities at the beginning of the pandemic (Diarmaid Hyland, 2021). While a majority of students reported having access to necessary study equipment and suitable study spaces during the pandemic, a minority cited economic or geographical constraints. Despite professors promptly providing extensive online resources, students encountered difficulties transitioning from traditional to remote learning methods. Over half of the respondents said that closures affected their capacity to learn mathematics through lectures, tutorials, and mathematics support services, while they also had trouble communicating and engaging with peers and hence low motivation. Moreover, the majority of students reported experiencing heightened levels of anxiety and isolation due to the university closures, with many proposing methods to enhance student interaction. The technical difficulties experienced by students during online assessments, particularly for high stakes assessments, raise significant concerns. Issues such as poor internet connectivity and hardware malfunctions highlight the impact of inequality on student outcomes, specifically within the online learning environment.

Despite the challenges a significant number of mathematics lecturers was satisfied with their pandemic assessment methods (Fhloinn, 2021). Teachers expressed their willingness to continue using some of the teaching practices used during Covid, planning to use a blend of the most effective components of face-to-face and remote teaching for future classes. Teaching methods such as the use of recorded lectures, more formative assessments such as online MCQs, instead of relying solely on the traditional summative closed book final exam, are a few of the approaches that educators would like to incorporate in their future practices. The integration of technology in teaching during the pandemic has equipped teachers with valuable skills and experience that can benefit them in the future.

2.3.3 Post Covid-19

Since the end of the pandemic, and the general transition of students and teachers moving back to in person teaching, assessment in mathematics has adapted from the way it was done during Covid. While the pandemic forced educators worldwide to teach differently, resulting in valuable skills, lessons and innovations, the extent to which traditional teaching methods will return post-pandemic remains uncertain. The shift to remote teaching during the pandemic has been a transformative experience for educators and students. As we navigate through this period, it's evident that remote teaching has the potential to continue even after the pandemic subsides. Remote teaching has opened doors for collaboration and innovation in curriculum design and delivery. It is crucial to reflect on the experiences of the pandemic and consider how to best incorporate the lessons learned into future educational practices (Johann Engelbrecht, 2023).

Gaps in the Literature

While there is a wide range of literature that is, at this time of writing, still coming out on the experiences of mathematics departments during Covid-19, there is a large gap in the research of what happened after (Henley & Hilliam, 2022; Irfan et al., 2020). As it stands, most of the research in the field focuses on the student's experiences and their motivations during the pandemic, as well as, tracking their performance (which was impacted by the change in teaching methods) (Bakar et al., 2022; Diarmaid Hyland, 2021; Ní Fhloinn & Fitzmaurice, 2021a). We find there is a gap in literature addressing the empirical data of assessment methods in the times after post Covid-19 era. Understanding the effects of the unprecedented pandemic on the methods of assessment of mathematics with quantitative research will aid understanding what the state of the art looks like in the field within higher education institutions.

Previously, as investigated by Iannone and Simpson (2022) we could infer a clear trend of how these assessments were changing and developing within the UK. With technology slowly becoming more of a focus on the field of research, there was a strong focus on whether the closed book exam would still have such a strong presence in the field of undergraduate mathematical assessment (Iannone & Simpson, 2011). However, with Covid-19 becoming a global issue, the natural progression of the methods used for assessment was met with a block that forced a change. We now investigate the long lasting effects, if any, these drastic changes starting in 2020 have had, filling in a gap in the literature in the field of undergraduate mathematics assessment within the scope of the UK.

Chapter 3

Methodology

3.1 Data Collection

The methods of the study reported in this paper closely resemble those used by Iannone and Simpson (2022).

In order to address the research questions, we carried out data collection of information about the assessment methods used in a number of different universities across the UK. The data compiled has been taken from publicly available sources online. This study considers university degree programs under BSc Mathematics, coded as 'G100' in the Higher Education Statistics Agency's JACS 3.0 system [1]. In particular data about each university's mathematical department was taken from each corresponding institution's online public website. Doing the data gathering this way allows us to have factual unbiased data that aligns with our research objectives. Note only data from summative assessment was considered, thus ignoring any information about formative methods.

3.1.1 Data Variables

The data gathered consisted of 5 main data variables. These were

(a) Course title

(b) Credits awarded

(c) Year typically taken in

(d) Compulsory/optional to curriculum

(e) Assessment type + weighting

In particular, assessment type + weighting was later broken up into more specific categories. The assessment type was broken up into 11 categories, which were decided based on the most recurring assessment types found during the data gathering. These are:

[1] https://www.hesa.ac.uk/support/documentation/jacs/jacs3-detailed

- **Closed book examination:** e.g. any examination where no support materials are available for students, taken in an enclosed examination hall with no outside aid. Only available resource could potentially be a calculator.

- **Open book examination:** e.g. any examination with any support materials available to the students, as well as, any online final examination.

- **Computer-based coursework:** e.g. any coursework involving programming or computer based activities, usually completed for a deadline in the student's own time.

- **Online quiz:** e.g. short set of online questions, typically between 10-15, performed under a set amount of time.

- **Written hand-ins/problem sheets:** e.g a marked worksheet completed in your own time, usually over the course of a week, based on material covered in class.

- **Individual project:** e.g. a substantial piece of written work, on a set topic or problem, undertaken over the course of a long period, such as a term or two. Written exclusively by one person.

- **Group project:** e.g. a substantial piece of written work, on a set topic or problem, undertaken over the course of a long period. Written exclusively by two or more people.

- **Class test:** e.g. any examinations held under normal class time in the tutorial/workshop environment.

- **Presentation:** e.g. oral, in person, presentation displaying a given topic, that is, any class material or results of a project.

- **Essay/Report:** e.g. extended piece of written work dealing with a class topic in depth.

- **Unspecified:** e.g. any other assessment method used, that was not specified in the information source.

Therefore, for each module we recorded the weight that each assessment method held towards the final mark for the class.

3.1.2 Data Sampling

The research involved non-probability sampling methods, in particular, the data sampling was reduced to universities in the United Kingdom (UK). Initially, the research considered all registered universities in Scotland that offered a BSc Mathematics degree qualification; of which any mathematics department that had sufficient information publicly available were studied. From this data gathering, a total of 10 Scottish universities were found that delivered a mathematics degree, out of which only seven universities had available data about assessment weighted averages and hence only those were treated (These are coded 1, 4, 5, 6, 7, 18, 20 - in order of appearance in the available league table - see table 4.1).

The scope of the research was then expanded to universities from the whole of the UK. Taking The Guardian's “The Best UK Universities 2024 – Ranking” [2] as a base, sampling was taken from the top, middle and bottom of this list. In this way, 6 universities were considered beginning from the top of the list, and going down; disregarding any that; (1) had already been studied because they were based in Scotland or (2) they didn't have the needed data available. In this same manner, a sample starting from the bottom of the list (and going up) was taken and a sample from the middle of the list, starting from 24 and going all the way down to 61. The positioning in the ranking was encoded from 1 to 24 in order of appearance in the league table. Hence we employed a hierarchical ordering instead of the given position in the ranking.

For each university, information from all modules available to students undertaking a 'G100' degree were taken. In order to focus our research, we only considered mathematics modules that were directly linked to the degree in hand; hence ignoring any information about modules students could take that were outside of the mathematics department for the given institution. Data from a total of 888 modules were collected, recording any details available about the assessment methods used, although note that not every department had these open to the public.

3.2 Data Handling

In this project, we have chosen Python for our data handling and analysis tasks due to its versatility and wide range of libraries, such as Pandas and NumPy, which are particularly efficient for handling large datasets. Additionally, Python's user-friendly syntax and extensive community support give it an edge over R.

To calculate an estimate of the percentage of marks accrued by closed book exams during the degree in one institution, we calculated the mean across all the modules on offer as part of the program, weighted according to the credits each module accrued.

Following the methodology used by Iannone and Simpson (2022), our modelling framework adopted an approach where it was assumed that one student would take all available modules offered within one mathematical department. The mean across all modules offered as part of the program was then calculated. Hence, to estimate the proportion of marks attributed to closed book exams within the degree program at a specific institution, we conducted a weighted calculation using the formula below 3.1. This involved multiplying each assessment method for each course by its corresponding credits (weights)

[2] https://www.theguardian.com/education/ng-interactive/2023/sep/09/the-guardian-university-guide-2024-the-rankings

and then dividing the sum of these products by the total number of credits, for each university across all years. This method provided an estimation of the overall contribution the closed book examination had to the cumulative marks earned throughout the program.

$$\text{Weighted Mean Average} = \frac{\sum_{i=0}^{n} x_i \cdot w_i}{\sum_{i=0}^{n} w_i} \tag{3.1}$$

All the 888 courses where further categorised into one of the 8 subject types:

- Pure mathematics
- Applied mathematics
- Statistics and probability
- Calculus
- History of mathematics and mathematics education
- Project
- Computer programming
- Skills

This was done manually and each course was only assigned one subject type to simplify analysis. The courses which fall under two or more subject types were assigned to one considering the overall distribution of the modules and to reduce any anomalies. This classification allowed for further analysis to be conducted at module level.

A comparison of the use of closed book exams as an assessment method across pre-honours and honours years was then conducted. Academic years were separated in this manner instead of addressing each year separately for a fair comparison with Scottish universities, which typically include four years. Thus, for our analysis, we combined years one and two of Scottish universities and considered them as the pre-honours years, alongside year one of English universities. For the honours years, we considered years two and three of English universities, and years three and four of Scottish universities.

Additionally, following the methods used in Iannone and Simpson (2022), our research involved comparing the use of closed book examination as the main assessment method for two distinct groups of universities: the Russell Group universities[3] (these are coded by 2, 3, 5, 6, 8, 9, 10, 12, and 21 by their position on the league table - see table 4.1) and Post-92 universities[4] (identified by codes 14, 15, 16, 22, 23, and 24 - see table 4.1), which is a common distinction for UK universities.

[3]Representing a selected group of 24 public research universities, institutions in the Russell Group are considered to be a leading universities in the UK.

[4]Post-92 universities include any former polytechnic that was given the university status after an education reform in 1992 in the UK.

Using the data provided by Iannone and Simpson (2022), a comparison was undertaken to assess the changes in assessment methods within UK mathematics university departments, just before the pandemic and in 2023. Notably, we deemed appropriate to include more types of module categories totaling 11 distinct assessment methods, in contrast to the 7 methods identified by Iannone and Simpson (2022). In the comparison we see that the assessments, 'online quiz', 'coding assessment, 'essay/report', were not categories in the 2019 dataset; while 'individual' and 'group' projects were merged together into one category to match our comparison data. Consequently, the comparison focused on the assessment method categories which were common to both datasets.

3.3 Edinburgh Case Study

In this section, we focus on the University of Edinburgh (UoE) as our primary institution and gather data exclusively from its online resources. Our objective is to detail the methods employed to address research question 7 (*See Chapter* 1). UoE was deliberately selected for this case study due to our access to a comprehensive database of module information, including detailed data on assessment methods, facilitated by our current enrollment at the university.

The data collection methodology closely follows the process outlined in section 3.1, maintaining consistency with the variables recorded as described in section 3.1.1. However, there is a variation in the method of data sampling. Here, our attention is directed towards all modules offered to students enrolled in the G100 BSc Mathematics program by the School of Mathematics at the University of Edinburgh. Consequently, a full sample was obtained in this instance. In order to facilitate both pre and post Covid analyses, apart from the current data on assessment, we now also collected the data of the University of Edinburgh sample correct at the time of the 2019/2020 diet (ie. pre-Covid data). This was available for us through the Degree Regulations and Programmes of Study (DRPS) resource provided by the UoE specifically. Note that for the general study with the larger UK sample (4), the pre-Covid Edinburgh data was not collected by our research group but taken from the original data set collected by Iannone and Simpson (2022).

Again, the methods explained in 3.2 were replicated when dealing with the data handling for the case study.

Note that following the pandemic, the School of Mathematics at UoE decided, as a default arrangement, to permit three A4 sheets of paper (equivalent to six pages) of notes for final exams, effectively classifying them as open book assessments [5]. Consequently, for the purpose of our data handling and categorization of assessment methods, we proceeded with the assumption that unless explicitly stated otherwise, all final examinations at UoE are open book.

[5] https://teaching.maths.ed.ac.uk/main/undergraduate/studies/assessment/examinations/sitting

Chapter 4

Data Analysis

4.1 Replication of Analysis from Previous Studies

Reflecting the approach detailed in Iannone and Simpson (2011) and Iannone and Simpson (2022), our analysis begins with calculating the weighted mean average for the percentage of CBEs at UK universities. Assigning credit values as weights to these percentages ensures that our calculations accurately represent each assessment's academic significance.

Figure 4.1 generated from Table 4.1 concisely illustrates the distribution of weighted average percentages for closed book examinations for pre-honours, honours and the overall degree. This visualisation clearly depicts the data trends and variances within each academic year. A one-way ANOVA was employed to further compare the means of CBEs across the two different academic years. A one-way ANOVA is suitable for comparing groups coming from the same population. The results from the test ($F = 0.0035, p = 0.9965$) suggest that the average use of closed-book examinations remains the same across different years as $p > 0.05$. Additionally, Figure 4.2 visualises the distribution of CBEs across the three university types - Russell, Post-92 and Other, demonstrating that Russell Group universities have a slightly higher mean weighted average of closed-book exams.

4.2 Overall Distribution of Assessments

Figure 4.3 represents the distribution of various assessment types used across UK universities' mathematics departments. According to Figure 4.3, CBEs are the most commonly employed assessment method, with an average usage of 57.80 %. This is followed by open book examinations at 9.87 %, which is considerably less but still the second most prevalent. The other assessment methods, such as coding assessments, online quizzes, and written handins/questions, are used to an even lesser extent, with average percentages of 0.53 %, 0.74 %, and 3.89 %, respectively.

Individual projects and group projects have moderate usage, with averages of 2.12% and 1.11%, respectively. Class Tests (2.35%), presentations (0.60%), and essays/reports (3.02%) are also part of the assessment mix but are less

Table 4.1: Weighted Average (%) of Closed Book Exams

Ranking	**University**	**Type**	**Pre-Honours**	**Honours**	**Total**
1	St Andrews	Other	68.00	67.95	67.96
2	UCL	Russell	77.50	84.00	82.79
3	Warwick	Russell	75.56	83.12	82.22
4	Aberdeen	Other	50.00	42.11	45.00
5	Glasgow	Russell	80.00	80.20	80.18
6	Edinburgh	Russell	0.00	0.00	0.00
7	Strathclyde	Other	65.71	77.50	72.34
8	Bristol	Russell	0.00	12.04	10.73
9	Exeter	Russell	75.00	63.00	65.53
10	York	Russell	65.00	85.00	81.84
11	Surrey	Other	65.62	69.52	68.72
12	Manchester	Russell	60.83	69.39	67.70
13	Essex	Other	0.00	18.82	12.80
14	Northumbria	Post92	75.00	64.44	67.08
15	Nottingham Trent	Post92	72.00	65.78	67.07
16	Liverpool John Moores	Post92	30.00	25.26	26.40
17	City	Other	77.00	72.68	73.79
18	Stirling	Other	60.00	35.88	47.13
19	Sussex	Other	78.75	63.12	66.25
20	Dundee	Other	40.00	66.23	54.95
21	Newcastle	Russell	75.83	75.61	75.66
22	Plymouth	Post92	60.00	40.45	44.64
23	London Met	Post92	42.50	33.53	36.40
24	Greenwich	Post92	36.25	20.40	24.24

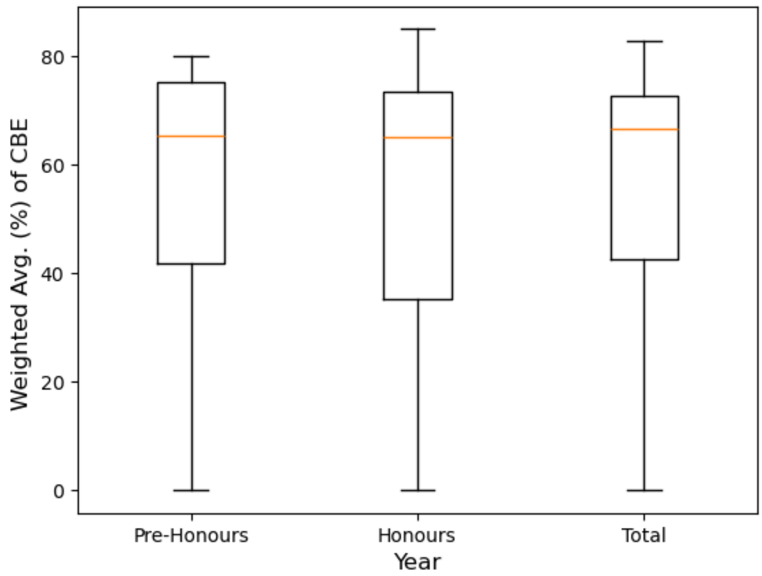

Figure 4.1: Distribution of Closed Book Examination by Year and in Total

common. A noteworthy portion of the assessments is categorised as 'Unspecified', accounting for 17.97 % of the average, indicating a significant proportion of assessments that are not classified into the standard types.

4.3 Analysis at Ranking Level

This section explores the relationship between assessment practices and the ranking of universities. We begin the analysis by calculating the Pearson coefficients between the weighted average of different assessment types and ranking. We then move to linear regression analysis between ranking (independent variable) and the weighted average of the assessment(dependent variable) with the strongest correlation coefficient of statistical significance.

4.3.1 Correlation Analysis

A Pearson correlation coefficient r of two variables, x and y, infers the linear correlation between them. This is calculated using the Equation 4.1 :

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \tag{4.1}$$

For our analysis, we take x to be assessments and y the university ranking. From Table 4.2, we can observe that the weighted average of overall examinations (Exam %) and the ranking of the university (Ranking) is strongly correlated with a correlation coefficient $r = -0.5465$. The negative sign indicates that universities with higher rankings, hence a lower number on the league

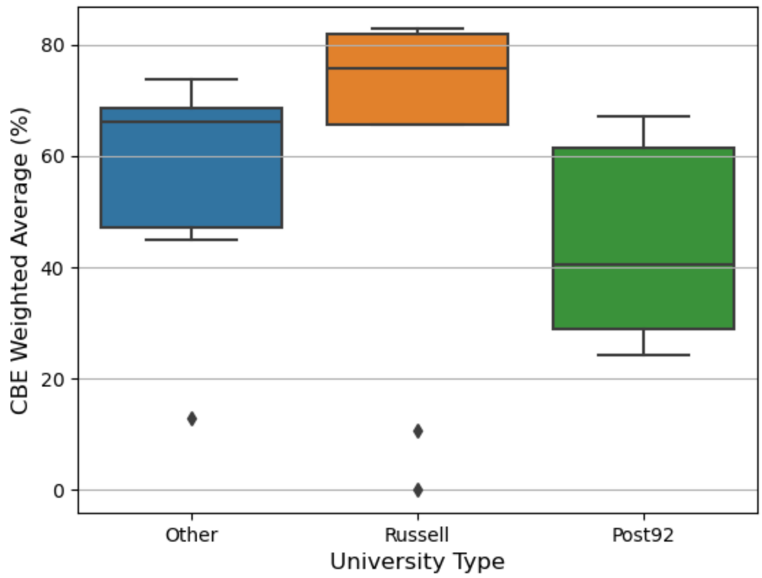

Figure 4.2: Distribution of Closed Book Examination by University Type

table, are more likely to use examinations to assess students. The p-value is 0.0057 (< 0.05), which indicates that this correlation is statistically significant.

Assessment	**r**	p
Exam%	-0.5465	0.0057
Presentation	-0.4161	0.0432
Coding assessment	-0.2834	0.1796
Online quiz	-0.2574	0.2247
CBE	-0.2241	0.2926
OBE	-0.1988	0.3517
Written handins/questions	-0.1468	0.4935
Group project	-0.0561	0.7945
Individual Project	0.0636	0.7678
Essay/ Report	0.1813	0.3965
Class Test	0.1948	0.3617
Unspecified	0.5308	0.0076
Coursework%	0.5465	0.0057

Table 4.2: Pearson Linear Coefficient and their p-value for Ranking and Assessment Weighted Averages

4.3.2 Linear Regression Analysis

From the previous section, we observed that the average use of examinations has a strong correlation with the ranking of a university. However, correlation does not imply causation. To further understand how these two variables

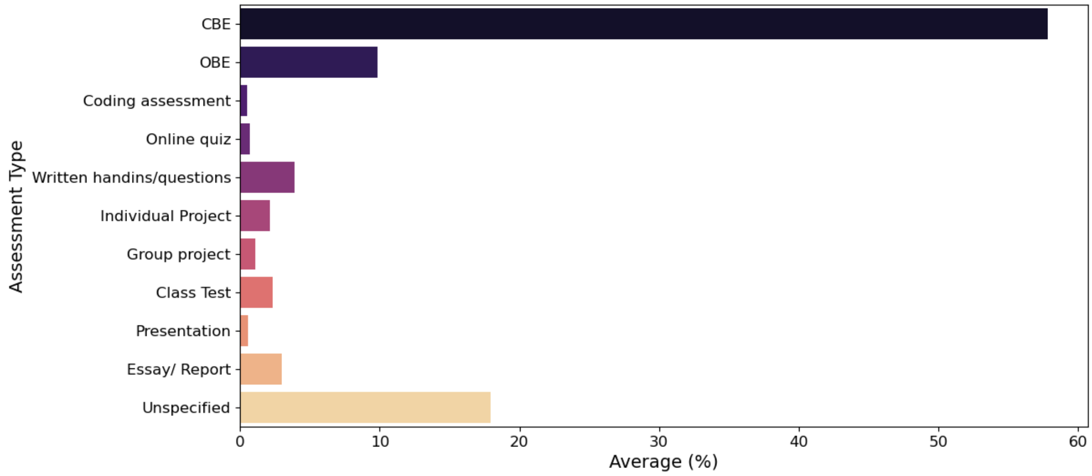

Figure 4.3: Average Use of Different Assessment Methods in UK Mathematics Departments

influence each other, we perform a linear regression analysis. The linear regression model offers a deeper look into this association through the equation:

$$y = \beta_0 + \beta_1 x \tag{4.2}$$

Here, let y (outcome variable) represent the weighted average of exams and x represent (predictor variable) represent the university ranking hierarchical position used to predict y. The results of the analysis are summarised in Table 4.3. The coefficient β_0 (80.1968) represents the value of y when $x = 0$. The coefficient β_1 (-1.3661) indicates the slope or the change in the outcome variable y with a unit change in predictor variable x in the positive direction. Thus, the equation becomes :

$$\text{Weighted Average of Exams} = 80.1968 - 1.3661 \cdot \text{University Ranking} \tag{4.3}$$

This model, depicted graphically with a line of best fit in the linear regression plot (Figure 4.4), demonstrates a clear negative trend, indicating that universities with higher rankings (with lower numeric values) have higher percentages of closed-book examinations.

The negative slope of the x (Ranking) coefficient in the regression model suggests a significant inverse relationship, as higher numerical value of university rankings are associated with a decrease in exam percentages. The p-value of 0.006 for this coefficient underlines its statistical significance. Furthermore, the F-statistic of 9.369 with a corresponding p-value of 0.00572 indicates that the model as a whole is statistically significant. Despite this, with an R-squared value of 0.299, the model accounts for approximately 29.99% of the variability in exam percentages, which suggests that factors other than university rankings also play a considerable role. They are summarised in Table 4.4.

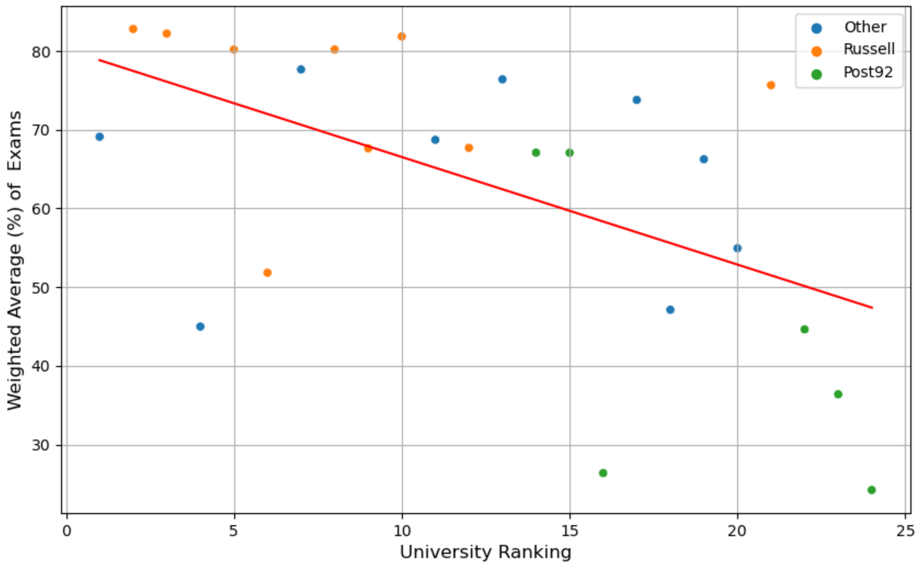

Figure 4.4: Predictive Model illustrating the Relationship between Ranking and Weighted Average Exam%

	coef	std err	t	$P > \|t\|$	[0.025	0.975]
const	80.1968	6.377	12.575	0.000	66.971	93.423
Ranking	-1.3661	0.446	-3.061	0.006	-2.292	-0.441

Table 4.3: Results from Linear Regression Analysis

4.4 Analysis Based on University Type

This section explores the distribution of assessment types across different categories of universities - Post-92, Russell and Other. It is then followed by the tests conducted to assess any statistically significant difference in assessment practices across these three groups.

4.4.1 Distribution of Assessments

Figure 4.5 illustrates the distribution of various assessment types used across the three categories of universities for mathematics degrees. Russell Group universities are observed to predominantly employ traditional examination formats, with CBEs and OBEs constituting 60.74 % and 13.70 % of assessments, respectively. This preference highlights a commitment to established methods for evaluating individual academic performance within structured environments.

In contrast, Post-92 institutions show a complete absence of OBEs (0.00%) and a reduced reliance on CBEs (44.31%).

Universities classified under ”Other” demonstrate a moderate engagement

Linear Regression Statistics	
R-squared:	0.299
Adj. R-squared:	0.267
F-statistic:	9.369
Prob (F-statistic):	0.00572

Table 4.4: Statistics from Linear Regression Analysis

with traditional assessments, recording a 56.55 % usage for CBEs and a modest engagement with OBEs at 7.79 %, positioning themselves in between Russell Group and Post-92 universities regarding their assessment strategies.

The analysis further reveals a nuanced adoption of non-traditional assessments such as coding assessments and online quizzes, particularly within Russell Group universities, where they account for 1.05% and 1.24%, respectively.

Additionally, written hand-ins/questions and individual projects are used by all three types of institutions. Russell Group universities exhibit a notable preference for written assessments (5.03%) and maintain a balanced approach to individual projects (2.37%).

While group projects and presentations are less common, their presence is notable, especially in Russell Group universities, with averages of 1.63% for group projects and 0.83% for presentations.

Lastly, a significant proportion of assessments within Post-92 universities falls under the unspecified category (40.93%), in stark contrast to the lower percentages observed in Russell (8.95%) and Other (17.00%) institutions. This considerable variance may indicate the use of a diverse range of innovative assessment methods that defy conventional categorisation or perhaps reflect a gap in information about assessment practices.

4.5 Analysis at Degree Course Level

The following analysis explores the assessment methods used across different academic years for our set of universities offering mathematics undergraduate degrees.

4.5.1 Observations

For the honours academic year, the data reveals that examinations account for 66.70% of the assessment strategy, with coursework making up the remaining 33.32%. This contrasts with the pre-honours year, where exams constitute 61.65% and coursework 38.35%. In terms of specific assessment types within these broader categories, CBEs are the most prominent in both years, comprising 55.94% in the honours year and 54.19% in the pre-honours year. OBEs also play a significant role, making up 10.72% and 7.45% of assessments in the honours and pre-honours years, respectively.

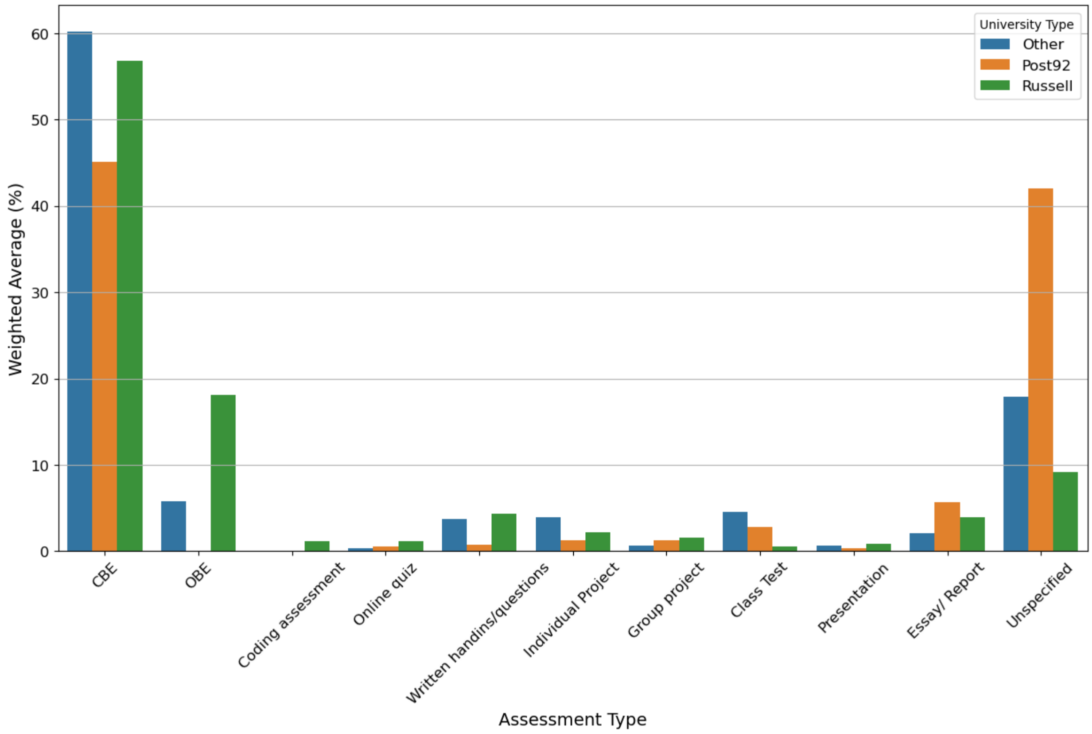

Figure 4.5: Comparison of Assessment Methods Utilization by University Type (Russell, Post-92, Other)

The utilisation of coding assessments and online quizzes is relatively low in both academic years. Written handins/questions and class tests are more common in the pre-honours year, accounting for 4.31% and 4.53%, as opposed to 3.08% and 1.79% in the honours year.

Individual and group projects are slightly more prevalent in the honours year, with percentages of 3.18% for individual projects and 1.17% for group projects, compared to 0.80% and 1.33% in the pre-honours year. The honours year also shows higher usage of presentations and essays/reports, at 0.82% and 4.32%, respectively, compared to 0.22% and 1.56% in the pre-honours year.

A notable portion of assessments is categorised as unspecified in both years, with 18.35% in the honours year and 22.06% in the pre-honours year. This indicates a variety of assessment methods are utilised that are not explicitly detailed in the provided categories.

4.5.2 Mann–Whitney U test

We applied the Mann-Whitney U test (*See Appendix* B) to solidify our comparison of the assessment practices between pre-honours and honours years and examine any statistically significant difference between them. The results of the test are in Table 4.5

The test results highlight a clear difference in the approach to assessment across these academic levels for certain assessment types while showing consistency for others.

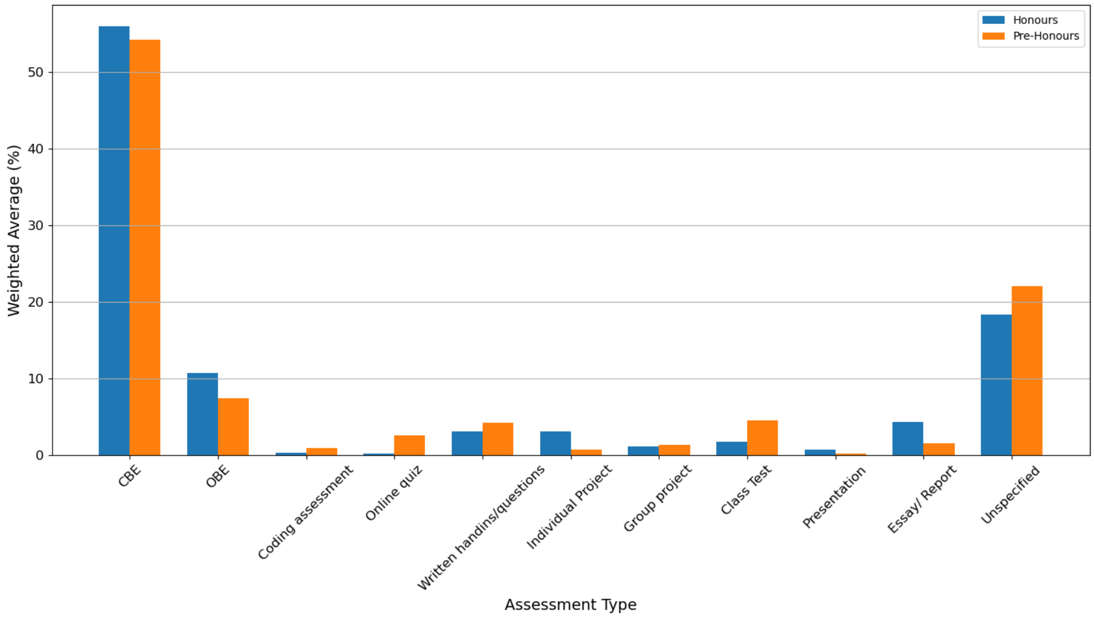

Figure 4.6: Comparison of Assessment Methods used in Degree Course Level

Significant differences ($p < 0.05$) are observed in the overall distribution of examination and coursework. Within specific assessment types, CBEs, online quizzes, class tests, individual projects and unspecified assessments show statistically significant differences. Their p-values suggest that the distribution and, implicitly, the reliance on these assessment methods vary markedly between the honours and pre-honours years.

Conversely, OBEs, coding assessments, written handins/questions, group projects, presentations, and essays/reports show no statistically significant differences, with p-values ranging from 0.12354 to 0.49138. This lack of significant difference suggests a consistent use of these assessment methods across both honours and pre-honours years.

4.6 Analysis at Module level

In this section, we delve into the assessment practices across different mathematics modules, aiming to understand the variety and commonality in assessment practices within different subject areas. We use principal component analysis and cluster analysis to uncover potential patterns in how different modules approach assessments.

4.6.1 Principal Component Analysis

Principal Component Analysis (PCA) (*See Appendix* A) was employed to address the complexity of the multidimensional data obtained from diverse assessment methods across various mathematics modules. The high dimensionality of the dataset posed a significant challenge for direct analysis and visualisation. PCA provided a solution by transforming the original variables

Assessment Type	U**-Statistic**	p**-value**
Exam%	82703.0	0.00000
Coursework%	46165.0	0.00000
CBE	75660.0	0.00011
OBE	66619.0	0.14703
Coding assessment	62938.0	0.12354
Online quiz	59786.5	0.00000
Written handins/questions	62773.5	0.48452
Individual Project	66433.5	0.02242
Group project	63520.0	0.49138
Class Test	57673.0	0.00004
Presentation	65544.5	0.12895
Essay/ Report	64985.5	0.45933
Unspecified	54930.5	0.00149

Table 4.5: Mann-Whitney U Test Results for Assessment Types: Pre-Honours vs Honours

into a new set of orthogonal variables known as principal components, which represent the directions of maximum variance. The first two principal components encapsulate key differences in assessment practices (PCA Component 1 and PCA Component 2, as illustrated in Figure 4.7), with PCA1 reflecting the broader methodological divergence and PCA2 capturing more nuanced variations. These were selected for further analysis as they capture the most substantial variation within the dataset, hence providing a two-dimensional approximation that simplifies the data's complexity and makes it easier to visualise while retaining its most informative aspects.

4.6.2 Cluster Analysis

Building on the PCA, cluster analysis enabled us to identify and visualise clusters of modules that share similar assessment characteristics. The K-means algorithm (*See Appendix* A) sorts modules into clusters when mapped onto the PCA framework. The graph (Figure 4.7) illustrates distinct groupings, with programming, project and history of mathematics + mathematics education modules demonstrating unique assessment patterns, while calculus, applied mathematics, and pure mathematics show remarkable similarity in their approach.

4.6.3 Inferences

From the clusters identified, it is clear that modules such as calculus, applied mathematics, and pure mathematics share a pedagogical approach to assessment, likely emphasising traditional problem-solving skills. On the other end of the spectrum, programming and project modules stand apart, suggesting a focus on practical application and synthesis of knowledge. The position of the

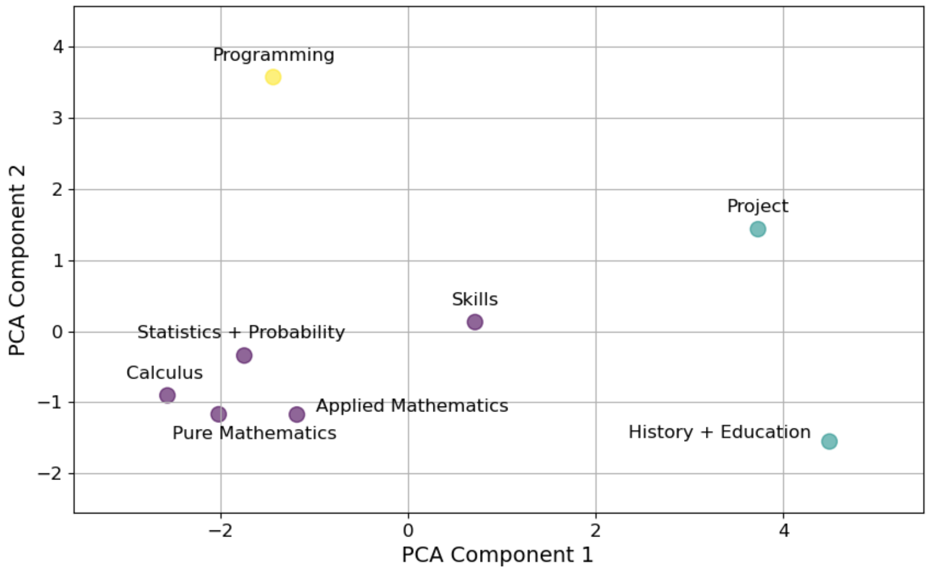

Figure 4.7: Cluster Analysis of Assessment Types for Different Modules of Mathematics

history + education module indicates a deviation from conventional assessment practices, possibly favouring written assignments and presentations. We find the skills module to be in between the analytical modules and theoretical modules, indicating a mix of both conventional assessment practices as well as essay based assessments.

4.7 Analysis at Regional Level

In this section, we explore the differences in assessment practices in Mathematics degrees between England and Scotland. This comparison is motivated by the distinct structural differences between English and Scottish universities, with the latter typically incorporating an additional fourth year of study. We aim to examine whether this structural variation influences the approaches employed in assessing students' academic performance.

4.7.1 Distribution of Assessments

Figure 4.8 illustrates the overall average distribution of Exams and Coursework between England and Scotland whereas Figure 4.9 visualises the distribution of specific assessment types.

In England, exams constitute 66.49% of the assessment weight, with coursework accounting for 33.53%. Within specific assessment types, CBEs account for 56.35%, while OBEs account for 10.09%. Other forms of assessment, including coding assessments (0.54%), online quizzes (0.37%), written handins/questions (2.99%), individual projects (2.05%), group projects (1.30%), class

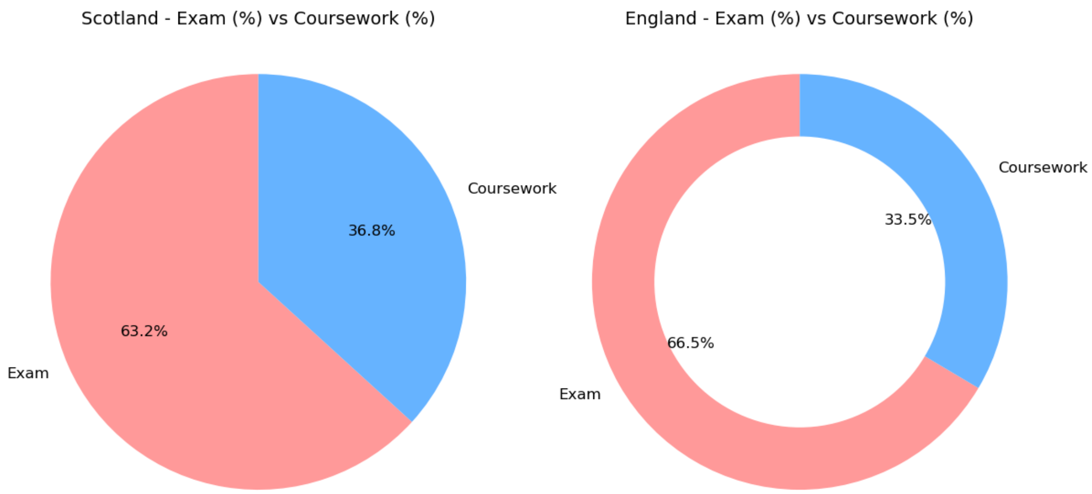

Figure 4.8: Weighted Average of Exam and Coursework Assessments in Scotland and England.

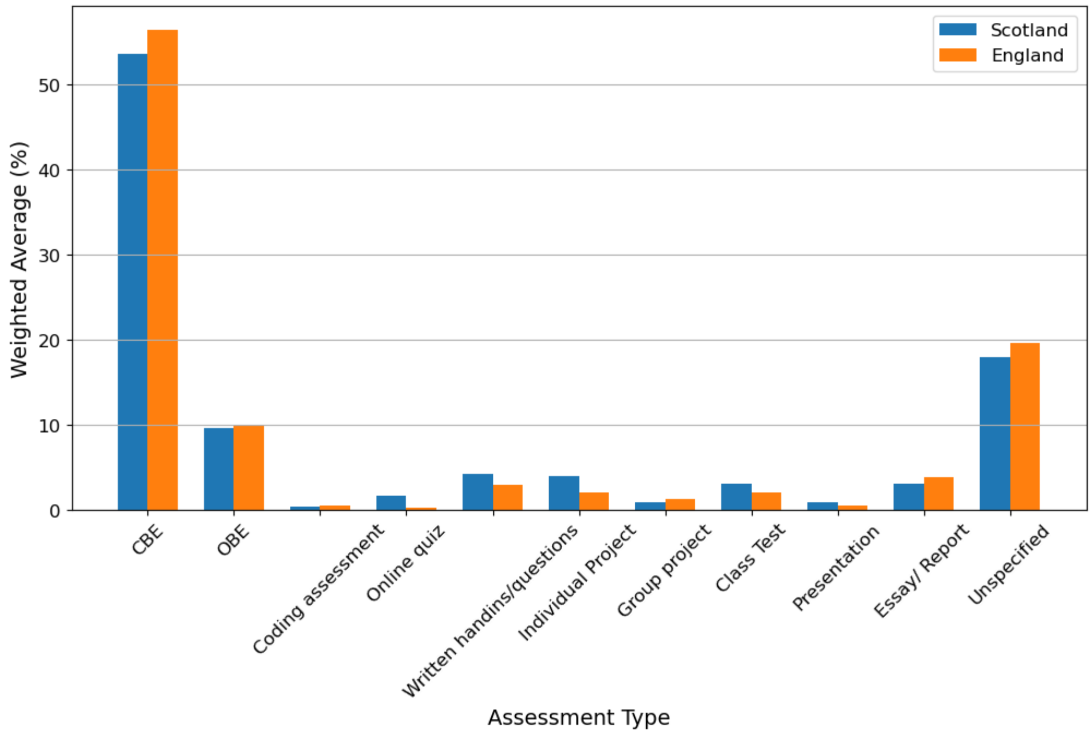

Figure 4.9: Distribution of Assessments in Mathematics Degrees : Scotland and England

tests (2.14%), presentations (0.58%), and essays/reports (3.90%), contribute to the diverse evaluation framework. A significant 19.68% of the assessments are unspecified, highlighting the varied nature of assessment strategies.

Conversely, Scotland places a slightly lower emphasis on examinations, making up 63.24% of the assessment weight, with coursework at 36.76%. CBEs in Scottish universities are utilized to a slightly lesser extent at 53.58%, while the use of OBEs is approximately at the same 9.66%. The approach to other assessment types in Scotland varies with written handins/questions (4.25%), individual projects (4.04%), essays/reports (3.18%), class tests (3.12%), online quizzes (1.75%) receiving more focus. Coding assessments (0.48%), group projects (0.98%), presentations (0.92%), and also play integral roles. Similar to England, a substantial portion of the assessments remains unspecified at 18.04%.

4.7.2 Significance Test : Mann-Whitney U test

The Mann-Whitney U test was chosen to compare the assessment practices in undergraduate mathematics degrees between England and Scotland due to its appropriateness to handle unequal sample sizes without making assumptions about the underlying data. The results of this test are summarised in Table 4.6.

Assessment Type	***p*-value**
CBE	0.159
OBE	0.060
Coding Assessment	0.904
Online Quiz	0.003
Written Handins/Questions	0.386
Individual Project	0.022
Group Project	0.855
Class Test	0.745
Presentation	0.502
Essay/ Report	0.652
Unspecified	0.519

Table 4.6: Mann-Whitney U Test Results for Assessment Types for England and Scotland

The results from the Mann-Whitney U tests reveal differences ($p < 0.05$) in the use of online quizzes (p-value = 0.003) and individual projects (p-value = 0.022) between the two regions, indicating that these assessment types are utilised differently in England compared to Scotland. For all other assessment types, including CBEs, OBEs, coding assessments, written handins/questions, group projects, class tests, presentation, essays/ reports, the p-values did not indicate significant differences, suggesting that these assessment practices are relatively consistent between the two regions.

4.8 Evolution of Mathematical Assessment Practices: 2019 to 2023

The weighted average use of various assessments before and after the pandemic demonstrates significant shifts in assessment practices due to the pandemic. Specifically, CBEs saw a marked decrease from 70.06% before the pandemic to 55.54% afterwards, indicating a substantial shift away from this traditional assessment method in the post-pandemic era. Unspecified assessments showed a negligible decrease from 19.44% to 19.20%, suggesting that the proportion of assessments not clearly categorised remained relatively stable. Conversely, the utilisation of written hand-ins/questions almost doubled, growing from 1.67% to 3.37%, which highlights an increased reliance on this form of assessment following the pandemic. Project-based assessments maintained a relatively stable role in assessment strategies across the period, shifting only slightly from 3.97% before the pandemic to 3.84% afterwards. The use of presentations decreased from 0.98% to 0.68%, possibly reflecting the challenges or adjustments in conducting presentations in a post-pandemic educational context. Class tests also saw a reduction in use, from 3.57% to 2.42%, suggesting a reduction in the use of this assessment due to the suspension of on-campus classes in light of the pandemic. Notably, OBEs witnessed a dramatic increase from a virtually negligible 0.05% before the pandemic to 9.97% afterwards, marking a significant shift towards open-book formats as a response to the pandemic.

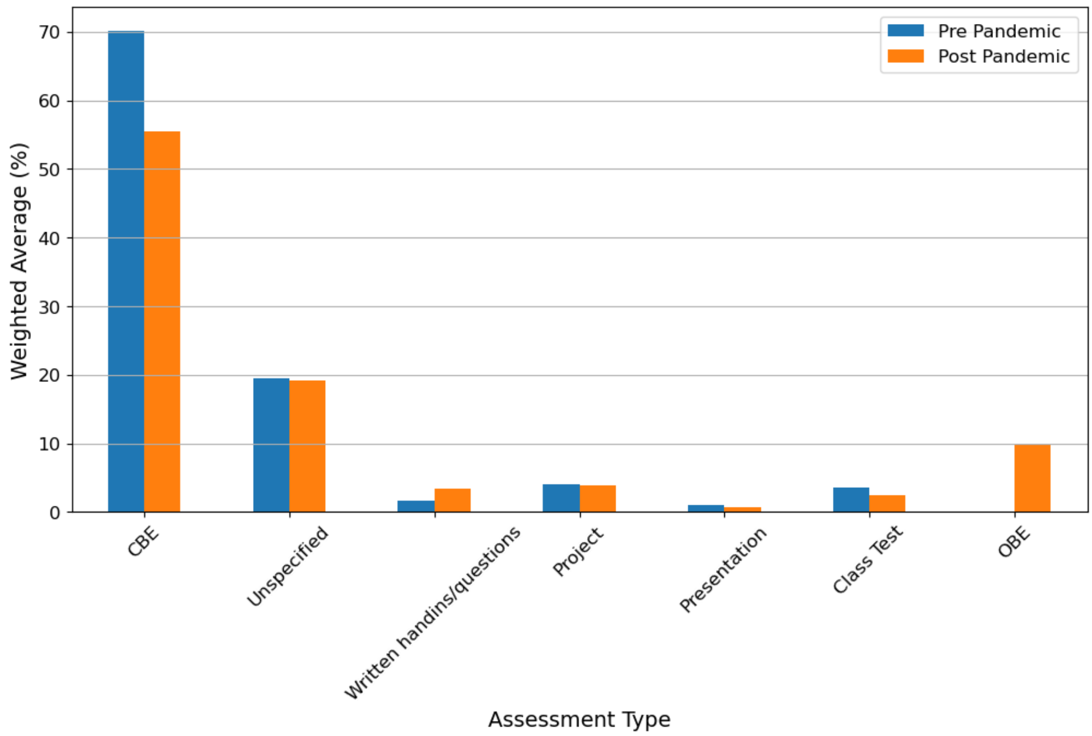

Figure 4.10: Comparison of Assessment Methods in Mathematics Courses : Pre-Pandemic vs Post-Pandemic

Figure 4.10 provides a visual representation of the changes in assessment methods over the years, highlighting the transition towards a more varied and integrative approach post-pandemic.

4.8.1 Paired t-test for pre and post pandemic comparison

A paired t-test (*See Appendix* B) was utilised to compare the weighted average use of various assessments before and after the pandemic. This test was chosen due to the nature of the data, where each assessment type's averages before and after the pandemic could be considered as paired observations. This test aims to determine if there is a statistically significant difference in the mean values of these pairs.

The results of the paired t-test indicated a t-statistic of approximately 0.248 with a p-value of 0.812. This suggests that there is no statistically significant difference in the weighted average use of various assessments before versus after the pandemic, indicating that the overall assessment practices did not significantly change due to the pandemic.

4.8.2 Mann-Whitney U Test for 2019 and 2023 Comparison

To perform a direct comparison of the distributions of individual assessment practices between the years 2019 and 2023, the Mann-Whitney U test was applied. This non-parametric test was chosen due to its suitability for comparing two independent samples without assuming normality in their distributions. It evaluates whether one of the two samples of independent observations tends to have larger values than the other.

Assessment Type	**U-Statistic**	**p-value**
CBE	624103.5	< 0.00001
OBE	429801.5	< 0.00001
Written Handins/Questions	453184.0	0.00001
Class Test	509320.5	0.00454
Presentation	491621.0	0.44580
Unspecified	475586.5	0.28508
Project	489050.0	0.91498

Table 4.7: Mann-Whitney U test results comparing 2019 and 2023

Table 4.7 shows statistically significant differences in the use of CBEs, OBEs, written handins/questions, and class tests between 2019 and 2023. These results indicate notable shifts in assessment practices for these types, highlighting changes in educational assessment strategies over the pandemic.

4.9 Validity of Statistical Tests and Potential Errors

The statistical tests conducted, including the paired t-test, Mann-Whitney *U* test, linear regression, and Pearson correlation, were selected on the basis data characteristics and research questions. The paired t-test was applied to compare the distribution of assessments in 2019 and 2023, with the assumption of normal distribution of differences verified through Shapiro-Wilk test (*See Appendix* B), thereby ensuring its appropriateness. The Mann-Whitney *U* test, used for comparing two independent samples, does not assume normal distribution, making it suitable as it does not make any assumptions about the underlying data.

Pearson correlation analysis was conducted to assess the linear relationship between university ranking and assessment types, providing insights into the direction and strength of these relationships. Linear regression analysis was employed to explore the relationship between university rankings and the weighted average of exams, with the regression model's validity supported by the R-squared value.

Despite careful selection of statistical tests, potential errors should be acknowledged. These include:

- **Type I and Type II Errors:** The possibility of falsely rejecting a true null hypothesis or failing to reject a false null hypothesis, respectively, inherent to statistical hypothesis testing.

- **Assumption Violations:** While efforts were made to validate assumptions (e.g., normality for the paired t-test), deviations in assumptions for other tests could impact their validity.

- **Sampling Errors:** The sampled data may not perfectly represent the population, potentially impacting the inferences drawn for the general population from this study.

- **Measurement Errors:** Inaccuracies in data collection or processing could introduce biases in the analysis.

In conclusion, the statistical tests applied are deemed valid for the purposes of this analysis, with considerations for their validity and potential errors carefully weighed. The findings contribute to our understanding of the relationships between assessment types, university rankings, academic years, regions, and the impact of the pandemic, albeit with an acknowledgement of the limitations.

Chapter 5

Edinburgh Case Study

This chapter analyzes the evolution of assessment practices at Edinburgh University from 2019 to 2023, contextualized by the pandemic's impact on education.

Reflecting on the shifts in assessment practices highlighted earlier, it's clear the pandemic has significantly impacted how the mathematics departments of universities in the United Kingdom evaluate student learning. To further explore these changes, we focus on the University of Edinburgh's School of Mathematics. Our choice of Edinburgh stems from our direct experience as students here, providing us with detailed insight into the assessment practices before and after the pandemic.

Currently, the School of Mathematics at the University of Edinburgh offers a diverse range of courses at the undergraduate level, totalling 49 courses across various departments. Of these, 18 courses fall under the purview of Pure Mathematics, demonstrating the school's strong emphasis on foundational mathematical theories and principles. Similarly, 5 courses are dedicated to Applied Mathematics, highlighting a commitment to applying mathematical concepts to solve practical problems in science, engineering etc. The School of Mathematics also offers 3 courses in Computational Mathematics and Programming and 9 courses in Statistics and Probability.

By examining these courses, we aim to uncover how assessment strategies within the department have evolved, reflecting broader trends in educational adaptation to unprecedented global challenges.

5.1 Distribution of Assessments

Table 5.1 shows the weighted average distribution of assessments for the University of Edinburgh's mathematics courses in 2023. Open book exams are the most prominent, accounting for 51.84% of assessments. The Essay/ Report category also plays a significant role, contributing 13.16% to the assessment mix. Other notable assessment types include Online Quizzes (7.72%), Written Hand-ins/Questions (4.85%), and Individual Projects (5.88%), with Coding Assessments (2.5%) and Group Projects (2.13%) also contributing to the

Table 5.1: Weighted Average Distribution of Assessments in 2023

Assessment Type	Weighted Average (%)
OBE	51.84
Essay/Report	13.16
Unspecified	9.56
Online quiz	7.72
Individual project	5.88
Written handins/questions	4.85
Coding assessment	2.50
Presentation	2.35
Group project	2.13
CBE	0.00
Class test	0.00

assessment mix. Presentations account for 2.35% of the assessment strategy. Both CBEs and Class Tests are absent, each marked at 0.00%, indicating a shift away from these traditional assessment formats. The 'Unspecified' category, previously more prominent, has been adjusted to 9.56%, reflecting a more precise categorization of assessment types in the updated analysis.

5.2 Assessment Distribution by Year

Table 5.2: Weighted Average Distribution of Assessments by Year

Assessment Type	Pre-Honours (%)	Honours (%)
OBE	35.0	57.9
Coding assessment	0.56	3.2
Online quiz	28.06	0.4
Written handins/questions	13.33	1.8
Individual Project	0.0	8.0
Group project	1.39	2.4
Essay/ Report	4.17	16.4
Presentation	0.0	3.2
Unspecified	17.49	6.7

Table 5.2 shows the weighted average distribution of assessments for pre-honours and honours years at the University of Edinburgh. In the pre-honours year, online quizzes (28.06%) and written hand-ins (13.33%) are significantly more prevalent than in the honours year, where they are less emphasized. In contrast, OBEs play a much larger role in the honours year, making up 57.9% of the weighted average. Individual projects are introduced in the honours year with an 8.0% weighting and are absent from the pre-honours year. This is possibly due the introduction to honours year projects. The essays and reports

category also sees an increase in the honours year to 16.4%, up from 4.17% in pre-honours. The unspecified assessments shows a decrease in the honours year to 6.7% from 17.49% in pre-honours, indicating a clearer definition of assessment types as students progress. Figure 5.1 visually represents the differentiation in assessment types between the two academic levels, with pre-honours favoring more frequent and varied forms of assessments and honours focusing significantly on OBE and individual research projects.

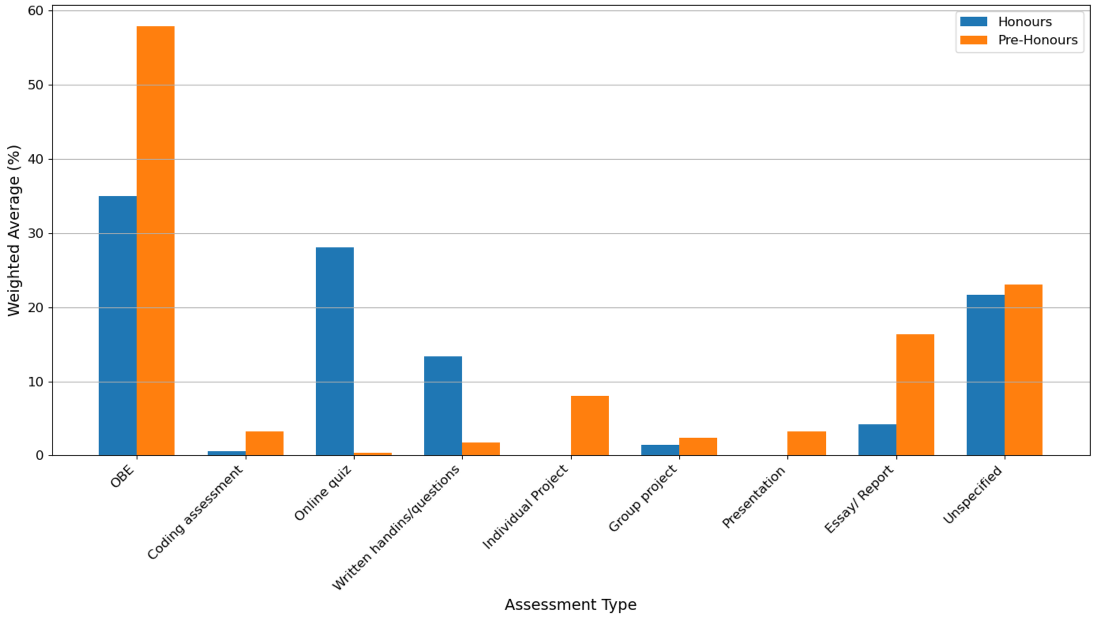

Figure 5.1: Distribution of Assessments in across Academic Years : Pre-Honours and Honours

5.2.1 Mann-Whitney U Test

The Mann-Whitney U test was chosen for further comparison of assessment utilisation for the two academic years as it is the best fit for comparing two independent samples (pre-honours and honours) when the data do not satisfy the normal distribution and equality of variances assumptions are required for other comparison tests.

Table 5.3 summarises the results from the Mann-Whitney U tests and reveals significant differences in the use of certain assessment types between the pre-honours and honours years, specifically in OBEs (p-value = 0.004) and online quizzes (p-value = 1.88×10^{-6}), indicating a statistically significant difference in their distribution. The written hand-ins/questions also showed a significant difference with a p-value of 0.006. These findings suggest variations in assessment strategies priorities between the two academic levels.

Conversely, assessment types such as CBE, coding assessments, individual projects, group projects, class tests, presentations and essays/reports did not show statistically significant differences, as evidenced by p-values well above

Table 5.3: Mann-Whitney U Test Results for Assessment Types : Pre-Honours vs Honours

Assessment Type	U Statistic	P-value
CBE	209.0	1.0000
OBE	92.0	0.0041
Coding assessment	204.5	0.8552
Online quiz	337.5	1.8816×10^{-6}
Written handins/questions	300.5	0.0062
Individual Project	203.5	0.6248
Group project	216.0	0.7078
Class Test	209.0	1.0000
Presentation	187.0	0.2778
Essay/ Report	200.0	0.7206
Unspecified	164.0	0.2585

the 0.05 threshold. This indicates a consistent application of these assessment methods across both pre-honours and honours years.

5.3 Assessment Comparison : 2019 vs 2023

5.3.1 Weighted Average Comparison

Figure 5.2 reflects the changes in assessment practices at the University of Edinburgh before and after the onset of the pandemic. In 2019, the exam component was predominant, making up 77.57% of assessments, in contrast to coursework at 20.49%. Post-pandemic, in 2023, the ratio of exams to coursework has evened out, with exams at 51.84% and coursework at 48.16%.

In terms of specific assessment types, 2019 was characterized by a significant reliance on CBEs at 67.86%, while other assessment methods such as online quizzes, class tests, and individual projects were used to a lesser extent. Notably, OBEs, coding assessments, written hand-ins/questions, group projects, presentations, and essays/reports were either not utilized or were negligible.

By 2023, the assessment landscape diversifies significantly. CBEs no longer hold a share of the assessments, while OBEs surge to 51.84%. Coding Assessments, online quizzes, written hand-ins/questions, individual projects, group projects, presentations, and essays/reports now all have a stake in the assessment distribution, with weights ranging from 2.13% to 13.16%. The 'Unspecified' category, which held a substantial portion in 2019, has decreased by 2023.

5.3.2 Mann-Whitney U Test Results

The results from Mann-Whitney U test shows statistically significant differences in several assessment types from 2019 to 2023 (Table 5.4) . Notably,

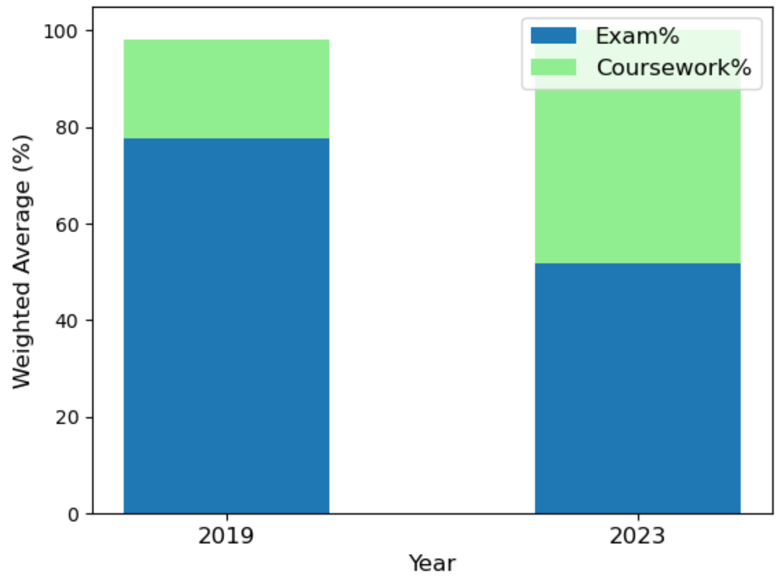

Figure 5.2: Comparison of Exam% and Coursework% between 2019 and 2023.

the p-values for CBEs and OBEs are significantly low ($p < 0.0001$), indicating major shifts in assessment strategy following the pandemic. Coding assessments, online quizzes, and essays/reports also show significant changes ($p < 0.05$) with p-values of 0.0038, 0.0011, and 0.0014, respectively. Written hand-ins/questions experienced substantial changes, evidenced by a p-value of less than 0.0001. The presence of group projects and presentations became more pronounced, as suggested by their p-values of 0.0263 and 0.0099. In contrast, individual projects and class tests remained relatively stable, with higher p-values suggesting no significant change in their usage.

5.4 Discussion

In light of our overall research it is important to note that in context, we are dealing with a Scottish university ranked as a Russell Group university. The University of Edinburgh has some details that might differ from other universities we used for our total data collection. It has no modules that are assessed entirely 100% with an exam. While grades rely significantly on final examination for most courses (usually 80% or even 95%), no module is offered that is assessed entirely 100% by final examination. Edinburgh University stands out for its extensive offering of Computational and Programming Courses compared to other universities in our dataset. Specifically, it provides 5 dedicated programming courses and incorporates computational modules into several statistical courses. Following the pandemic, the Department of Mathematics at the University of Edinburgh introduced a new policy mandating that all courses (unless otherwise chosen by the course organiser) allow for the use of three sheets of paper (equivalent to six pages) during final examinations by de-

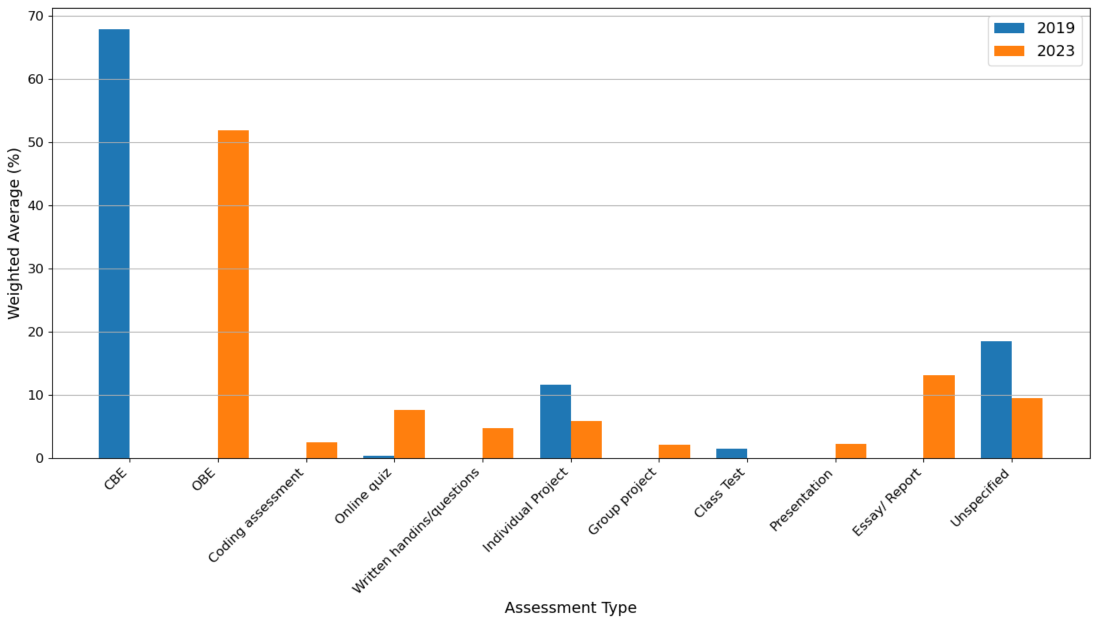

Figure 5.3: Bar chart of the weighted averages for various assessment types in 2019 and 2023

fault. This shift may signal a departure from the strictly closed-book examination format that was prevalent before the Covid-19 pandemic. The compulsory honours courses at the University of Edinburgh integrate skill-building components aimed at enhancing students' abilities, such as public speaking and presentation skills. For example, in the Honours Algebra course, 50% of the overall grade is allocated to developing presentation skills, requiring students to present on various topics.

After the pandemic, the University of Edinburgh has significantly increased its utilization of online quizzes for student assessment, particularly focusing on pre-honours years. Additionally, there has been a notable rise in the incorporation of coding assessments post-Covid. These changes suggest a potential transition towards a more blended learning approach, integrating methods adopted during the pandemic such as online quizzes and computational/coding assessments.

Furthermore, the shift to open book examinations might be a result of COVID-19, during which most exams were conducted as open book and open web. This trend reflects a move towards emphasizing student learning over memorization.

5.5 Conclusion

Focusing on the data from a single university unveils more substantial shifts in assessment practices pre and post Covid compared to the overall trends observed across the entire UK university sample. While the broader analysis indicated a persistent reliance on traditional exam-based assessments, Edinburgh's approach diverged significantly in the post-pandemic era. The focused

Table 5.4: Mann-Whitney U Test Results for Assessment Types: 2019 vs 2023

Assessment Type	U Statistic	p-value
CBE	3699.5	< 0.0001
OBE	480.0	< 0.0001
Coding assessment	1760.0	0.0038
Online quiz	1663.5	0.0011
Written handins/questions	1400.0	< 0.0001
Individual Project	1993.5	0.5938
Group project	1840.0	0.0263
Class Test	1984.5	0.4433
Presentation	1800.0	0.0099
Unspecified	2315.5	0.0760
Essay/ Report	1720.0	0.0014

analysis on this institution revealed a pronounced departure from closed-book examinations towards more innovative and varied assessment methods. Open-book exams emerged as the predominant form of evaluation, accompanied by the increased adoption of coding assessments, online quizzes, written assignments, projects, presentations, and essays.

Chapter 6

Discussion and Inference

6.1 Comparison with Previous Years

6.1.1 Overall comparison

As observed in Section 4.8, the main takeaway from the data analysis is that when comparing the assessment methods currently used in UK universities to the methods used in 2019, there is no statistically significant difference that we can report. Hence yet again, little has changed in the overall use of assessment methods. However, there exists an evident attempt to change individual assessment methods, most notably CBEs and OBEs.

The relationship with the groupings of the university types also remains the same. We notice that as in the previous research paper, Russell Group dominates the other institutions in the marks accrued by CBE, closely followed by the other non-classified universities, and Post-92 shows the lowest weighted average. However, when looking at the medians of the analysis, all of the medians have reduced slightly, compared to the 2019 data, where we see decreases of 75.7% to 77.6% and 71.9% to 66.25% in the dominating categories Russell Group and others, respectively. Notably, Post-92 universities, demonstrate a greater inclination for employing coursework assessment methods compared to universities originating from more traditional academic backgrounds, like Russell Group institutions, which typically lean towards CBEs. The lack of change in the results of this analysis demonstrates a commitment to the methods of different institutions and the deep-rooted traditions of these.

The distribution of the weighted mean percentage of marks accrued from CBEs across the different degree course levels remains with no change when compared to the pre-Covid data. The pattern remains where no significance has been found in the data, meaning that no statistical statement can be made about a difference between pre-honours and honours years.

While the overall preference for assessment methods may not have undergone significant changes, there is now a notable increase in the variety of methods employed. Despite this, closed-book exams remain the favoured choice across most departments. Examples of these expanded methods include computer-based coursework, group projects, online quizzes, and essays/reports. The integration of online assessments and modules may have been accelerated by the pandemic, which necessitated the testing of these approaches.

6.1.2 Comparison at module level

Probability and Statistics

Every department in our sample provided courses in probability and statistics, constituting approximately 13% of all courses offered. This percentage closely aligns with pre-pandemic data. Notably, 10% of the Statistics courses were evaluated through an open book exam. Given the extensive formulas involved in this subject, the provision of a formula sheet may classify these exams as open book assessments.

History of Mathematics and Mathematics Education

Out of 24 departments included in our sample, 9 offered a history of mathematics and/or mathematics education module, totaling 11 such modules across all departments. Out of the 11 modules, specifically 2 are history of mathematics modules, while the remaining are mathematics education modules. The predominant assessment method for this module among the majority of departments is entirely through coursework, with only 2 departments employing examinations as well.

When comparing to pre-pandemic data, a smaller percentage of the collected sample now offers these modules. Prior to the pandemic, over half of the departments provided mathematics education modules, whereas 4 out of the 23 departments offered history modules. This is nearly double compared to the current data.

Projects

Nearly every department in our sample, with just two exceptions, provides a final year project or dissertation module. These projects can be collaborative (6 out of 24 departments offered group or double projects) or individual and occasionally offer double or, in one department's case, triple credits. Typically, they focus on topics in pure mathematics and statistics and all departments in our sample offering this module schedule it for the final year of study, which corresponds to year 3 for English universities and year 4 for Scottish universities.

Compared to the pre-pandemic dataset, the number of departments offering a project module increased from around 60% to over 90% post-pandemic.

Computational and Programming

Our sample indicates that the percentage of programming and computational modules has remained consistent at around 5% compared to pre-pandemic data. However, four departments within our sample did not offer any computational modules. These modules typically involve computer-based projects or class tests, resulting in lower rates of closed-book examinations. Cluster analysis suggests that programming does not align with other modules in terms of assessment methods. This divergence is likely attributable to its highly computational nature, which makes traditional examination formats impractical compared to other modules.

Skills

Ten departments in our sample offer generic skills modules, primarily focusing on employability skills such as CV writing. This represents an increase from the seven departments observed in the pre-pandemic sample. Additionally, 20 departments offered mathematics skills modules covering various topics including communicating mathematics, analysis and proof writing. Compared to the pre-pandemic data, there has been an increase in the number of departments offering this module, rising from 78% to 83%. These modules are typically assessed entirely by coursework, predominantly in the form of essays.

6.2 Discussion

The most salient finding that our research shows is the emergence of the open book examination; which was not present in the 2022 paper (Iannone & Simpson, 2022). While the literature before the emergence of Covid reports on a student preference towards open book exams (Williams & Wong, 2009); much research during the pandemic showed the popular use of this method as an alternative method to the CBE (Cassibba et al., 2020; Henley & Hilliam, 2022). It is due to this connection to the appropriate literature that we can attribute the sudden rise of the use of the OBE in a non-pandemic context to the changes to the assessment diets during 2020. We can argue that the appearance of this new assessment method, in general, brings down the use of other methods, such as CBE, class tests, and presentations. In this sense, if more modules are now assessed with OBEs, as a direct consequence, there will be less modules assessed by alternative methods, which brings down the overall proportion of marks accrued by closed book examinations. In contrast, there is a statistically significant increase in the use of hand-ins and the percentage weight on these. We suggest this, as well as the decrease in class tests, are a direct consequence of the impact of the pandemic. While hand-ins can be associated to remote pieces of work, class test directly implies a need to be in classroom. Thus, we see assessment strategies linked directly to the pandemic that have carried on in the post-pandemic context.

As described above (*See* 6.1.2), nearly every department now includes a final year project, a significant increase from the 60% offering it in the pre-pandemic sample. The increased emphasis on these, may result from an heightened focus of employability, reported by Iannone and Simpson (2022). The inclusion of a project module serves multiple purposes, extending beyond traditional assessment of subject knowledge and testing different student's skills. These projects often require students to produce extensive reports or essays, refining their writing skills while explaining their findings. Moreover, students apply statistical tests learned during their coursework in practical analyses. Many universities incorporate a presentation component, evaluating students' communication and public speaking abilities. Additionally, the incorporation of group projects enhances students' teamwork, collaboration, leadership, organizational, and time management skills. Furthermore, the large increase in this could be seen as a result of an emphasis of expanding the field of mathematical assessment in order to demonstrate a closer relationship between research and practice (Nortvedt & Buchholtz, 2018). Projects are a direct application of applying mathematical concepts to the discipline, giving students the chance to apply their concepts to a situational study. Additionally, there has been a noticeable decrease in the number of departments offering mathematics education modules compared to pre-pandemic data. This decline may also be attributed to the integration of these modules as available topics for final year student projects. The effort to enhance employability and improve students' skills is also shown by the increase in departments offering skills modules, with nearly all departments now incorporating them into their curriculum. One might anticipate an increase in the number of computational/programming modules offered post-Covid due to their compatibility with remote learning and followed by the observed increase in the pre-pandemic the data from 2011 to 2020 by Iannone and Simpson (2022). Surprisingly, however, the percentage of computational modules offered remained unchanged compared to pre-pandemic levels.

In essence, what the results from the data analysis tell us is that there is still a high reliance on examinations as the main assessment method, whether this is closed or open book. We can identify precisely, the changes in the assessment strategy between pre-honours and honours years. The tests, ran in Section 4.5.2, demonstrate that during honours years students accrue more marks based on CBEs than in pre-honours years. This is a compelling result when contextualised into the overall degree grading system, as in the UK, the grade obtained in pre-honours years doesn't contribute to the final degree classification. Furthermore, the data also shows a relationship between the weighted average % of examinations and the ranking level, exhibiting a negative correlation. Hence, we can ascribe higher examination weighting to institutions rated higher in the rankings, understanding then, that it is highly regarded within a university's performance to measure more of student's grades with examinations. The interpretation we can deduce from these results, as well as the global distribution of assessments is that examinations are still the leading form of assessment within the field of mathematics. Even though some changes have been made, the idea that this is the best way to assess mathematics is still very respected; and agreed upon by many lecturers in the field who viewed it was the most efficient way to differentiate between stronger and weaker students (Iannone & Simpson, 2015a).

6.3 Limitations

An important limitation of our study involved the availability of data from university websites. As we tried to gather all the information we needed, we found that not every university website had the data we were looking for. Consequently, institutions lacking online data were excluded from our data collection process, potentially limiting the inclusivity of our findings. We noticed that universities lower in the chosen league table had less information available on assessment methods and hence we were not able to include as many as we wanted.

Specifically, within the context of Scottish universities, despite efforts to retrieve information regarding course offerings and assessment methods for all Scottish universities offering a mathematics degree, there were three universities with not sufficient data available online and hence we had to exclude them from our data gathering. Additionally, while our intent was to incorporate universities from across the UK into our data analysis, the absence of detailed online information from Welsh and Northern Irish universities led to their exclusion from our study. As a result, the research questions will only be answered in regards to universities from Scotland and England. Combining the data from Scottish and English universities for our analysis could present a limitation. This is because Scottish universities commonly offer four-year degree programs, whereas English universities typically offer three-year programs, resulting in a potential discrepancy in the duration of study years.

While some universities provided data on courses offered and general assessment methods, important details such as the nature of final exams (closed book or open book) or detailed information regarding coursework assessment types were frequently unavailable. As a result, all final examinations were considered to be closed book unless stated otherwise. This absence of specific assessment details limited the depth and completeness of our analysis.

Another drawback is the inability to explore the assessment distribution for courses spanning multiple mathematics department. This is particularly challenging due to the requirement of advanced data normalization techniques, a process that converts columns that contain nested data, in our case, the subject type, to the normal form (a new table). This task is complex and beyond the scope of our project. Consequently, this limitation hinders our ability to draw comprehensive comparisons of assessment standards between departments.

Note, as well, that the data was encoded as encountered on the available source, which might not always reflect the same practice between institutions (for example, where one institution might specify a percentage of the final grade is graded through a 'presentation', another might just have a presentation assessment encoded as a 'project').

We classified the universities in our sample according to the Guardian ranking; however, it is important to note that the rankings produced in this study, following the methods of Iannone and Simpson (2022), did not account for the actual differences or gaps in the scores between universities that dictated the ranking. This potentially could result in universities ranked closely together exhibiting significant differences in their underlying scores. Additionally, it's worth noting that the ranking positions utilized in the data collected in 2019 might differ entirely from those currently used and could include a distinct sample of universities. For our study, we employed the latest available rankings.

Chapter 7

Conclusion and Future Work

7.1 Concluding Remarks

In conclusion, the evidence suggests that mathematics remains a traditionalist subject, where the believe that a closed book exam is the most unbiased evaluation of a student's ability clearly endures. While there has been a diversification of assessment methods, closed book exams continue to be prevalent across mathematics departments, particularly among higher-ranking universities. Russell Group universities show a stronger inclination towards the use of a final examination to assess students compared to Post-92 institutions.

However, we do see the emergence of pockets of innovations that present new possibilities within the field of mathematics education assessment. Our data sample shows evidence of an evolution in assessment methods within mathematics departments in the UK since the Covid-19 pandemic, particularly with the establishment of progressive approaches, such open book examinations, into the assessment schemes, alongside traditional methods. This diversification in assessment strategies can be argued to have been stemmed, in part, from the demands imposed by the pandemic, prompting institutions to explore alternative evaluation modalities.

Looking at a comparative analysis between universities in England and Scotland, enforces the credibility of our UK-wide results, given that no significant difference can be found between the two countries. However, the focused analysis of one institution allowed to draw the attention on specific examples of the direct impact of the pandemic to the general assessment within a school of mathematics, for example, all exams being considered open book, unless specified otherwise.

7.2 Recommendations for Future Work

While surveys on student satisfaction (Kuznetsova, 2019) already exist, we recommend expanding this research by incorporating feedback specifically on the incorporation of open book examinations. This addition will provide valuable insights into students' perceptions and experiences with this evolving assessment format.

Furthermore, although this paper focuses primarily on higher education in the United Kingdom, we advocate extending the scope of research to include other nations. By comparing and contrasting practices and outcomes across different educational systems, we can gain a more comprehensive understanding of the implications of varying assessment methods on student success.

Bibliography

Bakar, S. A., Hadi, N. A., AlWadood, Z., & Yahya, A. A. (2022). Analysis of students performance in mathematics before and during covid-19 pandemic using pagerank: A preliminary study. *Asian Journal of Assessment in Teaching and Learning*, *12*(2), 100–109. https://doi.org/10.37134/ajatel.vol12.2.9.2022

Cassibba, R., Ferrarello, D., Mammana, M. F., Musso, P., Pennisi, M., & Taranto, E. (2020). Teaching mathematics at distance: A challenge for universities. *Education Sciences*, *11*(1), 1. https://doi.org/10.3390/educsci11010001

Chan, C. K. Y. (2023). A review of the changes in higher education assessment and grading policy during covid-19. *Assessment & Evaluation in Higher Education*, *48*(6), 874–887. https://doi.org/10.1080/02602938.2022.2140780

Chan, C. K., & Luk, L. Y. (2022). Eight years after the 3-3-4 curriculum reform: The current state of undergraduates' holistic competency development in hong kong. *Studies in Educational Evaluation*, *74*, 101168. https://doi.org/10.1016/j.stueduc.2022.101168

Crawford, J., Butler-Henderson, K., Rudolph, J., Malkawi, B., Glowatz, M., Burton, R., Magni, P. A., & Lam, S. (2020). Covid-19: 20 countries' higher education intra-period digital pedagogy responses. *Journal of Applied Learning & Teaching*, *3*(1), 1–20. https://doi.org/10.37074/jalt.2020.3.1.7

Diarmaid Hyland, A. O. (2021). The student perspective on teaching and assessment during initial covid-19 related closures at irish universities: Implications for the future. *Teaching Mathematics and its Applications: An International Journal of the IMA*, *40*, 455–477. https://doi.org/10.1093/teamat/hrab017

Eabhnat Ní Fhloinn, O. F. (2021). How and why? Technology and practices used by university mathematics lecturers for emergency remote teaching during the COVID-19 pandemic. *Teaching Mathematics and its Applications*, 392–416. https://doi.org/10.1093/teamat/hrab018

Fhloinn, O. F. E. N. (2021). Alternative mathematics assessment during university closures due to covid-19. *Irish Educational Studies*, 187–195. https://doi.org/10.1080/03323315.2021.1916556

Freedman, D., Pisani, R., & Purves, R. (2007). *Statistics* (International Student Edition, Fourth Edition). W.W. Norton & Co.

Henley, M., Matthew; Grove, & Hilliam, R. (2022). University Mathematics Assessment Practices During the Covid-19 Pandemic. *MSOR Connections*, 5–17. https://doi.org/10.21100/msor.v20i3.1325

Houston, K. (2001). Assessing undergraduate mathematics students. In *The teaching and learning of mathematics at university level: An icmi study* (pp. 407–422). Springer. https://doi.org/10.1007/0-306-47231-7_36

Iannone, P., & Simpson, A. (2011). The summative assessment diet: How we assess in mathematics degrees. *Teaching Mathematics and its Applications*, *30*(4), 186–196. https://doi.org/10.1093/teamat/hrr017

Iannone, P., & Simpson, A. (2015a). Mathematics lecturers' views of examinations: Tensions and possible resolutions. *Teaching Mathematics and Its Applications: International Journal of the IMA*, *34*(2), 71–82. https://doi.org/10.1093/teamat/hru024

Iannone, P., & Simpson, A. (2015b). Students' preferences in undergraduate mathematics assessment. *Studies in Higher Education*, *40*(6), 1046–1067. https://doi.org/10.1080/03075079.2013.858683

Iannone, P., & Simpson, A. (2022). How we assess mathematics degrees: The summative assessment diet a decade on. *Teaching Mathematics and its Applications*, *41*(1), 22–31. https://doi.org/10.1093/teamat/hrab007

Irfan, M., Kusumaningrum, B., Yulia, Y., & Widodo, S. A. (2020). Challenges during the pandemic: Use of e-learning in mathematics learning in higher education. *Infinity Journal*, *9*(2), 147–158. https://doi.org/10.22460/infinity.v9i2.p147-158

Johann Engelbrecht, M. C. B. G. K. (2023). Will we ever teach mathematics again in the way we used to before the pandemic? *ZDM Mathematics Education*, 1–16. https://doi.org/10.1007/s11858-022-01460-5

Kulm, G. (2013). Back to the future: Reclaiming effective mathematics assessment strategies. *Middle Grades Research Journal*, *8*(2), 1–10.

Kuznetsova, E. (2019). Evaluation and interpretation of student satisfaction with the quality of the university educational program in applied mathematics. *Teaching Mathematics and its Applications: An International Journal of the IMA*, *38*(2), 107–119. https://doi.org/10.1093/teamat/hry005

Laursen, S. L., & Rasmussen, C. (2019). I on the prize: Inquiry approaches in undergraduate mathematics. *International Journal of Research in Undergraduate Mathematics Education*, *5*, 129–146. https://doi.org/10.1007/s40753-019-00085-6

Ní Fhloinn, E., & Fitzmaurice, O. (2021a). Challenges and opportunities: Experiences of mathematics lecturers engaged in emergency remote teaching during the covid-19 pandemic. *Mathematics*, *9*(18), 2303. https://doi.org/doi.org/10.3390/math9182303

Ní Fhloinn, E., & Fitzmaurice., O. (2021b). Experiences of mathematics lecturers engaged in emergency remote teaching during the covid-19 pandemic. *Mathematics*. https://doi.org/10.3390/math9182303

Nortvedt, G. A., & Buchholtz, N. (2018). Assessment in mathematics education: Responding to issues regarding methodology, policy, and equity. *ZDM*, *50*(4), 555–570. https://doi.org/10.1007/s11858-018-0963-z

Panero, M., & Aldon, G. (2016). How teachers evolve their formative assessment practices when digital tools are involved in the classroom. *Digit Exp Math Educ*, *2*, 70–86. https://doi.org/10.1007/s40751-016-0012-x

QAA. (2023). *Subject benchmark statement: Mathematics, statistics and operational research*. https://www.qaa.ac.uk/docs/qaa/sbs/sbs-mathematics-statistics-and-operational-research-23.pdf?sfvrsn=5c71a881_12 (accessed: 11.02.2024).

Sangwin, C. (2004). Assessing mathematics automatically using computer algebra and the internet. *Teaching Mathematics and its Applications: An International Journal of the IMA*, *23*, 1–14. https://doi.org/10.1093/teamat/23.1.1

Sangwin, C. (2019a). Developing and evaluating an online linear algebra examination for university mathematics. *Eleventh Congress of the European Society for Research in Mathematics Education*, 4121–4128. https://hal.science/hal-02430556

Sangwin, C. (2015). Computer aided assessment of mathematics using stack. *Selected Regular Lectures from the 12th International Congress on Mathematical Education*, 695–713. https://doi.org/10.1007/978-3-319-17187-6

Sangwin, C. (2019b). Developing and evaluating an online linear algebra examination for university mathematics. *Eleventh Congress of the European Society for Research in Mathematics Education, Utrecht University*. https://hal.science/hal-02430556

Shorter, N. A., & Young, C. Y. (2011). Comparing assessment methods as predictors of student learning in an undergraduate mathematics course. *International Journal of Mathematical Education in Science and Technology*, *42*(8), 1061–1067. https://doi.org/10.1080/0020739X.2010.550946

Smith, G., & Wood, L. (2000). Assessment of learning in university mathematics. *International Journal of Mathematicla Education in Science and Technology*, *31*(1), 125–132. https://doi.org/10.1080/002073900287444

Steen, L. A. (2006). *Supporting assessment in undergraduate mathematics.* The Mathematical Association of America.

Trenholm, S., & Peschke, J. (2020). Teaching undergraduate mathematics fully online: A review from the perspective of communities of practice. *International Journal of Educational Technology in Higher Education*, 37. https://doi.org/10.1186/s41239-020-00215-0

Trevisan, O., De Rossi, M., & Grion, V. (2021). The positive in the tragic: Covid pandemic as an impetus for change in teaching and assessment in higher education. *Research on Education and Media*, *12*(1), 69–76.

Videnovic, M. (2017). Oral vs. written exams: What are we assessing in mathematics? https://doi.org/10.7251/OMEN1701001V

Voskoglou, M. (2019). Comparing teaching methods of mathematics at university level. *Education Sciences*, *9*(3), 204. https://doi.org/10.3390/educsci9030204

WHO. (2023). *Coronavirus disease (covid-19)*. Retrieved 2023, from https://www.who.int/news-room/fact-sheets/detail/coronavirus-disease-(covid-19) (accessed: 5.03.2024).

Williams, J. B., & Wong, A. (2009). The efficacy of final examinations: A comparative study of closed-book, invigilated exams and open-book, open-web exams. *40*(2), 227–236. https://doi.org/10.1111/j.1467-8535.2008.00929.x

Appendix A

Data Analysis Algorithms

Algorithm 1 Linear Regression Algorithm

function LINEARREGRESSION(X, Y)
  $n \leftarrow$ length of Y
  $\bar{x} \leftarrow$ mean of X
  $\bar{y} \leftarrow$ mean of Y
  $SS_{xy} \leftarrow \sum(x_i - \bar{x})(y_i - \bar{y})$
  $SS_{xx} \leftarrow \sum(x_i - \bar{x})^2$
  $\beta_1 \leftarrow SS_{xy}/SS_{xx}$
  $\beta_0 \leftarrow \bar{y} - \beta_1\bar{x}$
  repeat
    for each i from 1 to n **do**
      $y_{\text{pred},i} \leftarrow \beta_0 + \beta_1 x_i$
    end for
    $MSE \leftarrow \frac{1}{n}\sum(y_i - y_{\text{pred},i})^2$
  until MSE does not significantly decrease or until convergence
  return β_0, β_1
end function

Algorithm 2 K-means Clustering Algorithm

```
function KMEANSCLUSTERING(X, K)
    C ← select K initial centroids from X
    repeat
        S_i ← empty list for each i = 1, ..., K
        for each x_j in X do
            i ← arg min_i ||x_j − c_i||^2
            Append x_j to S_i
        end for
        for i = 1, ..., K do
            c_i ← mean of points in S_i
        end for
    until centroids do not change
    return C
end function
```

Algorithm 3 Principal Component Analysis (PCA) Algorithm

```
function PCA(X)
    m ← number of observations in X
    x̄ ← mean of each column of X
    B ← X − x̄ (mean centering the data)
    C ← 1/(m−1) B^T B (computing the covariance matrix)
    Compute eigenvalues and eigenvectors of C
    Sort eigenvectors by decreasing eigenvalues
    Choose k eigenvectors corresponding to the k largest eigenvalues to
  form W
    T ← BW (transforming the data into the new subspace)
    return T, W
end function
```

Appendix B

Statistical Tests

The statistical tests used for this project were taken from the book Freedman et al., 2007.

B.1 One-Way ANOVA

One-Way Analysis of Variance (ANOVA) is a statistical test used to determine if there are any significant differences between the means of three or more unrelated groups. It extends the t-test to more than two groups. The rationale behind ANOVA is to compare the variance between the groups with the variance within the groups.

Assumptions
The application of one-way ANOVA relies on several assumptions:

- Normal distribution of the dependent variable within each group.
- Independence of observations.
- Homogeneity of variances across the groups.

Hypotheses
ANOVA tests two hypotheses:

- Null hypothesis (H_0): The means of all groups are equal.
- Alternative hypothesis (H_1): At least one group mean is different from the others.

Test Statistic
The F-statistic is the ratio of the variance calculated between the groups to the variance within the groups. A higher F-value indicates a greater disparity between the group means relative to the variation within the groups, suggesting that at least one group mean significantly differs from the others.

Decision Rule
If the calculated F-statistic is greater than the critical value from the F-distribution

table at a specified significance level, the null hypothesis is rejected. This suggests significant differences exist among the group means.

B.2 Paired T-Test

The paired t-test is a statistical procedure used to determine whether the mean difference between two sets of observations is zero. It is applied when the observations are collected in pairs, or when each observation in one group is matched with a corresponding observation in the other group. This test is particularly useful for before-and-after measurements or when subjects are matched in pairs.

Assumptions
The paired t-test is based on several key assumptions:

- The differences between pairs are normally distributed.
- The pairs are selected randomly and are independent.

Hypotheses
The hypotheses for a paired t-test are formulated as follows:

- Null hypothesis (H_0): There is no difference in the mean of the paired differences; $\mu_d = 0$.
- Alternative hypothesis (H_1): There is a difference in the mean of the paired differences; $\mu_d \neq 0$.

Test Statistic
The test statistic for a paired t-test is calculated by dividing the mean difference between the paired observations by the standard error of the mean difference. The formula is given as:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

where $\bar{d}$ is the mean of the differences, s_d is the standard deviation of the differences, and n is the number of pairs.

Decision Rule
The decision to reject or fail to reject the null hypothesis is based on the calculated t-value and the critical t-value from the t-distribution table, considering the chosen significance level and the degrees of freedom (df = n - 1). If the absolute value of the calculated t exceeds the critical value, the null hypothesis is rejected, indicating a significant difference in means between the paired observations.

B.2.1 Mann-Whitney U Test

The Mann-Whitney U test, also known as the Wilcoxon rank-sum test, is a non-parametric statistical test used to compare differences between two independent samples. It is employed when the assumptions of the t-test are not met, specifically regarding the normality of the data or when dealing with ordinal data. The Mann-Whitney U test evaluates whether there is a difference in the distribution or medians of two groups.

Assumptions

Key assumptions of the Mann-Whitney U test include:

- The samples are independent.
- The response variable is at least ordinal.
- The distributions of the response variable for the groups are similar in shape.

Hypotheses

The hypotheses for the Mann-Whitney U test typically are:

- Null hypothesis (H_0): There is no difference in the distribution between the two groups.
- Alternative hypothesis (H_1): There is a difference in the distribution between the two groups.

Test Statistic

The Mann-Whitney U test involves ranking all the observations across both groups together, then calculating the U statistic for each group, which essentially measures the number of times observations in one group precede observations in the other group in the ranking. The test statistic used is the smaller of the two U values obtained.

Decision Rule

The decision on the null hypothesis is made by comparing the calculated U statistic to the critical value from the Mann-Whitney U distribution tables, based on the sample sizes of the two groups and the chosen significance level. If the calculated U is less than or equal to the critical value, the null hypothesis is rejected, indicating a statistically significant difference in the distributions between the two groups.

B.2.2 Shapiro-Wilk Test

The Shapiro-Wilk test is a widely used method for testing the normality of a dataset. It is particularly suitable for small sample sizes, typically less than 50 observations, but can also be applied to larger datasets. The Shapiro-Wilk test checks the null hypothesis that a sample x1, ..., xn came from a normally distributed population.

Assumptions

The Shapiro-Wilk test has the following assumptions:

- The sample is a simple random sample from its population.
- Each value in the sample is paired with a normal quantile.

Hypotheses

The hypotheses for the Shapiro-Wilk test are defined as:

- Null hypothesis (H_0): The population is normally distributed.
- Alternative hypothesis (H_1): The population is not normally distributed.

Test Statistic

The Shapiro-Wilk test statistic, W, is based on the correlation between the data and the corresponding normal scores. The test statistic is a ratio comparing the observed variance of sample order statistics to the expected variance under normality.

Decision Rule

To decide whether to reject the null hypothesis, we compare the value of the test statistic W to a critical value from the Shapiro-Wilk distribution. If W is less than the critical value for the chosen level of significance, we reject the null hypothesis, suggesting evidence against the normality of the data.