MSOR Connections

Articles, case studies and opinion pieces relating to innovative learning, teaching, assessment and support in Mathematics, Statistics and Operational Research in higher education.

Volume 23 No. 2.

THIS PAGE
DELIBERATELY
LEFT BLANK

Contents

For information about how to submit an article, notifications of new issues and further information relating to *MSOR Connections*, please visit https://journals.gre.ac.uk/index.php/msor.

This journal is published with the support of the **sigma** network and the University of Greenwich Faculty of Engineering and Science.

Editorial

Tony Mann, School of Computing and Mathematical Sciences, University of Greenwich, UK
Email: a.mann@gre.ac.uk

Welcome to the second issue of *MSOR Connections* for the academic year 2024/25. As usual the contents reflect current issues for the mathematics education and mathematics support communities. Dass and Price explore the differing quantitative training needs of research staff and postgraduate students. Three papers follow which relate to Artificial Intelligence (AI). Derrick and Weir explore how AI can be used to enhance statistics support for students; Chongchitnan, Parker, Mahal, and Petrie discuss the implications of generative AI for assessment in mathematics; and Falconer presents analysis of students' use of AI in history of mathematics assignments.

Finally we include a report by Bradshaw and myself of a workshop held in July 2024 which examined how to design assessments which support students' wellbeing. I am grateful to Alun Owen who oversaw the editorial process for this workshop report so that it could appear in a timely fashion.

As always, we are grateful to the authors for their contributions which we are sure readers will find useful, insightful, and sometimes provocative.

The next issue of *MSOR Connections* (Volume 23 No 3) will be a special issue comprising papers presented at the 2024 CETL-MSOR Conference, held at the University of Limerick in August 2024.

We are delighted to welcome three new editors for *MSOR Connections*. Anthony Cronin, Claire Ketnor, and Susan Pawley are very welcome additions to the editorial team and will be working on future issues of the journal. We are very grateful to the journal's Editorial Board and its Chair Ciarán Mac an Bhaird for overseeing the appointment of these new editors. We would also like to thank Peter Rowlett, who has stepped down as editor, for his invaluable contribution to the journal over many years.

Over the last few months the journal has been moved to an updated online platform, which will provide a much better experience for readers, authors, and reviewers. The update was overseen and implemented by Dave Puplett and Liam Clancy of the University of Greenwich, for whose assistance we are very grateful.

MSOR Connections can only function if the community it serves continues to provide content, so we strongly encourage you to consider writing research articles or case studies about your practice, accounts of your research into teaching, learning, assessment and support, and your opinions on issues you face in your work.

Another important way readers can help with the functioning of the journal is by volunteering as peer reviewers. When you register with the journal website, there is an option to tick to register as a reviewer. It is very helpful if you provide appropriate information in the 'reviewing interests' box, so that when we are selecting reviewers for a paper we can know what sorts of articles you feel comfortable reviewing. To submit an article or register as a reviewer, just go to http://journals.gre.ac.uk/ and look for *MSOR Connections*.

THIS PAGE
DELIBERATELY
LEFT BLANK

RESEARCH ARTICLE

Staff and Postgraduate Research Student Training Needs in Quantitative Methods: The Coventry Perspective

Yamuna Dass, **sigma**, Coventry University, Coventry, UK. Email: ab3390@coventry.ac.uk
Charlotte Price, **sigma**, Coventry University, Coventry, UK. Email: ad5778@coventry.ac.uk

Abstract

This paper explores the quantitative training needs of Postgraduate Researchers (PGRs) and university academic staff. An online survey was conducted by **sigma**, Coventry University's Mathematics and Statistics Support Service, to capture the perceptions and preferences of Coventry University PGRs and research staff around the quantitative training needed to support their research. Key topics of interest include the perceived need for training in specific statistical techniques, understanding statistical outputs and statistical software. The review suggests differences in the needs of PGRs and staff, with PGRs seeking foundational skills and staff requesting more advanced training. Additionally, staff with supervisory responsibilities emphasised the importance of PGRs developing skills in experimental design, data organisation, coding, analysis interpretation and presentation of findings - areas not mentioned by the PGRs. The findings also indicate that January and February are the most favoured months for training, with a significant preference for online delivery across participants. Furthermore, the review highlights the need for tailored workshops to address the diverse requirements of early stage researchers and experienced staff. Recommendations are provided, along with a description of changes implemented at Coventry University to better equip PGRs and staff with essential quantitative skills for their academic and professional careers.

Keywords: Quantitative training, statistical skills, Postgraduate Researchers (PGRs), staff members

1. Introduction

Every year, the Statistics Advisory Service team from **sigma**, Coventry University's Mathematics and Statistics Support Service, delivers a programme of statistics workshops for Postgraduate Researchers (PGRs) and staff members. The evolving nature of research increasingly requires skills in quantitative methods, even within disciplines traditionally dominated by a qualitative paradigm. This shift places significant pressure on PGRs and staff to develop quantitative skills. As such, in **sigma** we aim to provide quantitative research training to enhance researchers' skills and prepare them for careers that require them (ESRC, 2022; Vitae, 2011).

Before the Covid-19 pandemic in 2020, **sigma** offered a series of statistics workshops that were delivered in-person. Since then, they have been offered in-person or online, with the team tending to deliver software-related sessions in-person and theory-based sessions online. Historically, these workshops have been offered twice a year, in October/November and repeated in May/June.

Currently, the workshops cover a range of statistics techniques for conducting quantitative research projects. This includes questionnaire design, descriptive statistics, getting started using SPSS, t-tests, ANOVA, correlation and regression. Since the experience of the team is that many attendees have low confidence with regard to quantitative methods, the first few workshops are designed to cover basic concepts. Later workshops in the series move on to more specific techniques such as one-way ANOVA and linear regression.

Additionally, in recent years we have seen a rise in the number of PGRs and staff expressing an interest in learning to use a range of statistical software packages for their research work, which leads to demand for such software training. As such, we currently offer training workshops in both SPSS and R, with the latter increasing in popularity over time.

The content, structure and timing of the **sigma** workshops have not been reviewed for several years, while the demand for statistics support has continued to grow (Lawson et al., 2019). As more disciplines focus on data-driven research, having strong quantitative skills is becoming crucial in both academic and professional settings (British Academy, 2012; British Academy, 2015). As such, this paper aims to explore the training requirements in quantitative methods of PGRs and staff at Coventry University and seeks to provide insights to inform the planning of research methods workshops.

The objectives were to:

1. Explore PGR and staff perceptions of their training needs in quantitative research methods;
2. Develop an understanding of the quantitative research skills PGRs and staff may require for their own research work;
3. Capture insights from PGRs and staff on their workshop delivery preferences (e.g. timing and format of workshops);
4. Obtain an understanding from supervisors around the research skills they feel their PGRs need to develop.

2. Methods

1.1. Research design and data collection

To explore the views of PGRs and staff members in relation to their training needs in research methods, a survey was conducted using JISC Online Surveys (https://www.onlinesurveys.ac.uk/). The survey aimed to capture basic participant characteristics, perceptions of training needs in quantitative research methods and views on necessary skills required for undertaking research work. Participants were also asked about software usage, preferred training delivery mode (in-person or online) and preferred timing of training sessions during the year. Staff members with supervision responsibilities were asked further questions to capture their views on the necessary knowledge and skills for their PGRs. The survey questions can be found in Appendix A.

At Coventry University, research activity is concentrated within sixteen Research Centres. As such, the survey was promoted in-house through these Centres, as well as through the University's Doctoral College (DC) and within **sigma.** The survey link was distributed via each Centre's mailing list and newsletter, featured in the DC newsletter and posted on the **sigma** website. It was also emailed to PGRs and staff who had previously accessed **sigma**'s statistics support. Moreover, the link was shared on a staff mailing list targeting those interested in statistics, quantitative methods and/or research methodologies. The survey was conducted between October 2022 and January 2023.

1.2. Statistical analysis

Descriptive statistics, including frequencies with percentages, were obtained to provide an overview of participant characteristics and their training needs. Simple and clustered bar charts were used to visualise responses and chi-squared tests were used to explore associations, where relevant (Field, 2018). The analysis was undertaken using IBM SPSS Statistics 28 and Excel 365.

A basic content analysis was carried out on text responses to identify keywords (Krippendorff, 2018), as well as a basic thematic analysis to identify themes and explore participants' views about their training needs (Braun and Clarke, 2006).

Ethics approval for this study was provided by the Coventry University Research Ethics Committee (ref: P167064).

3. Results

There were 88 responses; 48 (54.5%) PGRs, 4 (4.5%) staff members who are also PGRs and 36 (41.0%) academic staff members. For this paper, the term “PGRs” will be used to represent both PGRs and staff members who are also PGRs.

A total of 80 participants belonged to a Research Centre, with nearly half of those from a social science background (n=36, 45.0%). This is perhaps not surprising given the nature of those disciplines and the likelihood of engaging with research methods. For example, health-related and education researchers were prevalent in the sample with just under a third of those aligned to research centres with a focus on health or biosciences (n=26, 32.5%) and 10 participants (12.5%) from the Centre for Global Learning, which addresses key educational challenges through research on global education and society. Among the PGRs (n=52), 20 were in their first year (38.5%), 14 in their second year (26.9%), 11 in their third year (21.2%) and 7 (13.4%) were at a later stage of their research programme. Additionally, 39 PGRs were engaged in full-time study (75.0%).

3.1. Perception of knowledge requirements for quantitative methods

Participants were asked to specify the level of knowledge they felt they needed in quantitative research methods (Appendix A, Q3), with almost all indicating the need for at least some knowledge (84/88). The responses were fairly evenly split: 25 specified a need for basic knowledge (28.4%), 33 indicated good working knowledge (37.5%) and 26 sought advanced knowledge (29.5%). Figure 1 contrasts the responses between PGRs and staff members, suggesting that staff members believe they require more advanced knowledge than PGRs. A chi-squared test of independence between required knowledge (basic, working, advanced) and role (PGR, staff) provided some support for the suggested association, with a p-value just above the 5% significance level $\chi^2(2, n=84) = 5.339$, $p=0.069$.

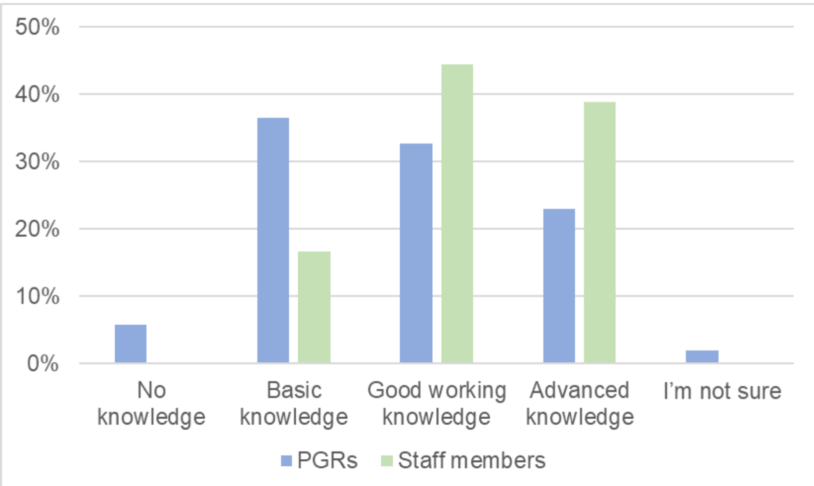

Figure 1: Level of knowledge around quantitative methods participants perceive they need by type of role; PGRs (n=52) and Staff members (n=36)

3.2. Quantitative skills participants felt they may need to develop

The development needs of participants were further explored to identify specific skills required for quantitative research methods (Appendix A, Q4 and Q5). Overall 75% of staff members (27/36) and 75% of PGRs (39/52) felt they may need to enhance their quantitative skills for their research work and/or postgraduate studies, suggesting agreement across the two groups in this respect. These participants (n=66) were then presented with a list of statistical skills and asked to select those they felt were necessary for them to learn.

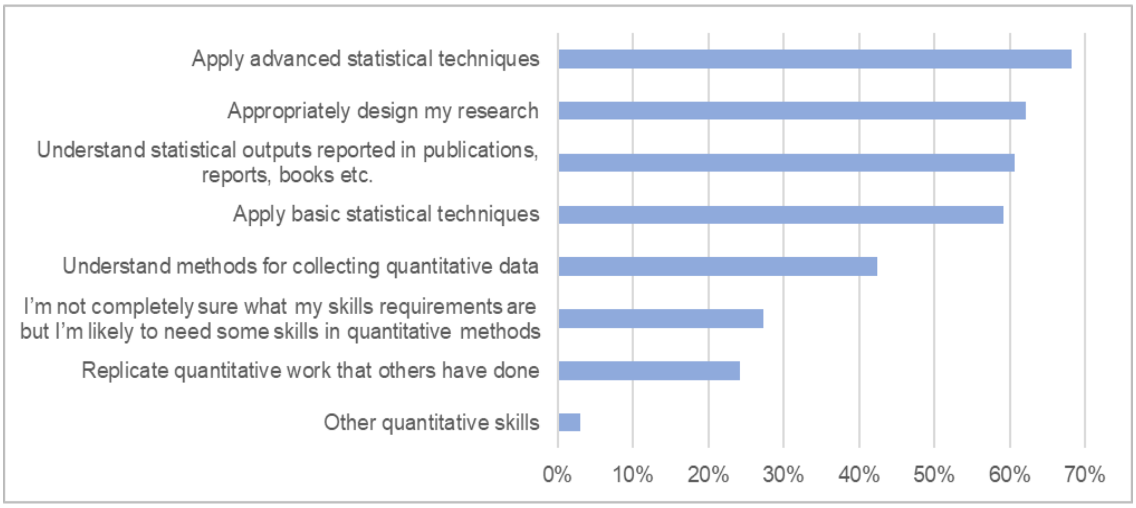

Figure 2: Quantitative skills participants felt that may need to develop (n=66)

From Figure 2, it is evident that many respondents feel that they would like to improve their skills in advanced statistical methods. It is also insightful to note there is a large demand for a better understanding of quantitative outputs in published sources, an important skill across the research spectrum, particularly with the increasing emphasis on undertaking systematic reviews.

Despite identifying a need for quantitative skills, nearly a third of respondents (18/66, 27.3%) were unsure about their specific developmental needs. Among these, the majority (16/18, 88.9%) felt they needed some level of knowledge to carry out quantitative research (5 basic, 8 good working and 3 advanced). This suggests a willingness to develop their skills in this area but highlights a need for additional guidance.

Participants' views on developing their quantitative skills were further analysed based on their roles, using chi-squared tests to explore associations between the type of role and each statistical skill. The results are presented in Table 1.

Table 1: Quantitative skills participants felt they may need to develop by type of role; PGRs (n=39) and Staff members (n=27)

Quantitative skills	PGRs (n=39)	Staff members (n=27)	Test statistic (χ^2)	p-value
Apply advanced statistical techniques; *n (%)*	25 (64.1)	20 (74.1)	0.731	0.392
Appropriately design my research; *n (%)*	28 (71.8)	13 (48.1)	3.791	0.052
Understand statistical outputs reported in publications, reports, books etc.; *n (%)*	28 (71.8)	12 (44.1)	4.999	0.025
Apply basic statistical techniques; *n (%)*	29 (74.4)	10 (37.0)	9.193	0.002
Understand methods for collecting quantitative data; *n (%)*	22 (56.4)	6 (22.2)	7.634	0.006
I'm not completely sure what my skills requirements are but I'm likely to need some skills in quantitative methods; *n (%)*	12 (30.8)	6 (22.2)	0.588	0.443
Replicate quantitative work that others have done; *n (%)*	10 (25.6)	6 (22.2)	0.102	0.750

PGRs appear more likely to want to develop a range of quantitative research skills compared to the staff members. This is likely due to the PGRs being at the beginning of their research journey compared to staff members. As shown in Table 1, there was evidence that PGRs were more likely to specify development needs for four of the skills compared to staff, namely appropriately designing their research (p=0.052), understanding statistical outputs reported in published sources (p=0.025), applying basic statistical techniques (p=0.002) and understanding different methodologies for collecting quantitative data (p=0.006).

These findings suggest that PGRs, at least in the early stages of their research programmes, need broad training to acquire a range of skills including statistical design, data collection, basic analysis and understanding reported statistical outputs. In this review, almost two-thirds of the PGRs were in the early stages of their research (i.e. year 1 or 2; 34/52, 65.4%), highlighting the importance of receiving training at the outset when planning their research project.

Overall, it appears that the training needs of PGRs and staff are different. Staff members indicate they would like to develop knowledge around more advanced statistical methods. This aligns with section 3.1 as PGRs primarily indicated a need for basic skills while staff required more advanced knowledge.

3.3. *Statistical techniques participants felt they may need to develop*

In addition to identifying particular areas for development, participants were asked to specify statistical techniques they felt they needed to know more about and/or use in their research work (Appendix A, Q6). Table 2 lists the more commonly mentioned techniques, indicating how frequently each one was mentioned. Techniques that were mentioned only once are listed in Appendix B.

Table 2: Statistical techniques participants felt they might need to know about and/or use in their research work (n=66)

Statistical technique	**Frequency**
Regression analysis	16
ANOVA	12
Descriptive statistics	9
Confidence intervals	9
T-tests	8
Parametric tests and non-parametric tests	7
Correlation tests	6
Checking statistical assumptions and dealing with violations	5
All tests	5
Logistic regression/binary logistic regression	4
Moderation analysis	4
Mediation analysis	3
MANOVA / MANCOVA	3
Bayesian statistics	3
Structural Equation Modelling	3
Use of R	3
Power analysis	3
Calculating effect size	2
Significance testing / p-value	2
Chi-square tests	2
Cluster analysis	2
Panel data	2
Advanced tests / Complex statistical modelling	2
I don't know	*21*

Techniques such as regression and ANOVA were most frequently mentioned. However, nearly one-third of the participants were unsure about which statistical methods they needed to learn for their research work (21/66, 31.8%). This group included 12 PGRs (year 1; n=5, year 2: n=5, year 3: n=1 and year 4, n=1) and 9 staff members. Nonetheless, the fourth-year PGR and a few uncertain staff members did mention specific advanced statistical techniques, such as probability distributions, Bayesian statistics, moderation and mediation regression. Overall, staff members demonstrated a greater interest in advanced statistical methods compared to the PGRs, reflecting the trends observed in Table 1, section 3.2.

A few staff also indicated that they would like to access a range of training opportunities and refresh their knowledge of basic skills as needed. One staff member noted that *"all [statistical techniques] would be helpful or at least have the option to access support/training on a vast array of techniques".*

Additionally, comments were made regarding the challenge of identifying training needs prior to starting a research project. For instance, one staff member mentioned that this *"depends on the scenario/project, making it difficult to predict the need before the project/need arises"*.

PGRs indicated an interest in acquiring both basic and advanced statistical techniques. This included descriptive statistics, t-tests, ANOVA, correlation, regression as well as more complex methods such as panel data analysis, moderation and mediation regression. Additionally, they expressed a need for support during the planning phase of their research and in selecting the appropriate statistical tests for their work. For example, one PGR commented they would like to know more about *"…making a data analysis plan/ how to get started with your research and analysis as a PhD student, and what you might need to know or plan for in advance to be prepared and not overwhelmed"*. Another PGR highlighted the importance of *"… identifying what else I can do with my data"*. This was echoed by a different PGR later in the findings (Appendix A, Q15), who requested *"support in gaining clarity around what is needed in the results section early on in the process… so it is clear from the outset"*.

Both staff and PGRs were keen to deepen their understanding of essential statistical techniques for undertaking research projects. Specifically, they highlighted the need for greater proficiency in interpreting p-values, performing power analyses, addressing violations of statistical assumptions and calculating effect sizes. This interest outlines the value of incorporating training on these statistical techniques into the offering, as both groups seemed keen to improve their skills in these areas.

3.4. Likelihood of attending quantitative skills workshops

Additionally, participants were surveyed on their likelihood of attending various quantitative skills workshops (Appendix A, Q7). Responses were grouped as likely to attend (very likely or likely), not likely to attend (very unlikely or unlikely) and unsure. Figure 3 suggests a strong interest in attending workshops on a range of quantitative research skills, including the use of statistical software packages. Many participants also expressed a desire to attend workshops to enhance their knowledge of advanced statistical techniques, whereas attending workshops on questionnaire design was less popular.

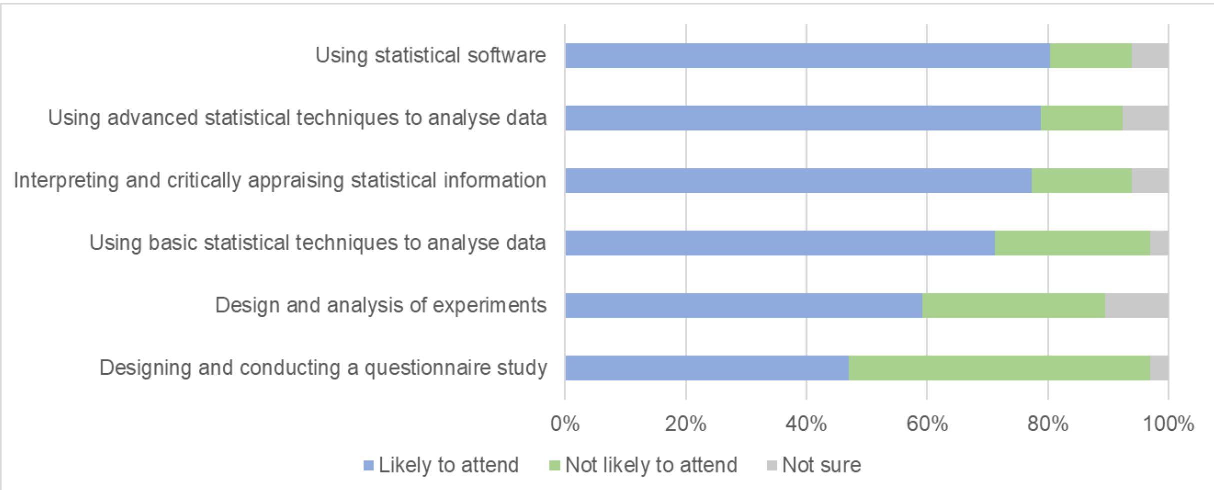

Figure 3: Likelihood of attending each quantitative skills workshop if offered (n=66)

When comparing the likelihood of attending workshops between PGRs and staff, those who were unsure were excluded from the analysis due to the small sample sizes. As a result, the responses relating to each workshop are variable. Chi-squared tests were conducted to explore associations between role and likelihood of attending each quantitative skills workshop (not likely or likely).

Table 3: Participants who were likely to attend each quantitative skills workshop by type of role; PGRs and Staff members

Quantitative skills workshop	PGRs	Staff members	Test statistic (χ^2)	p-value
Using advanced statistical techniques to analyse data; n (%)	30 (81.1)	22 (91.7)	1.297	0.255
Using statistical software; n (%)	33 (89.2)	20 (80.0)	1.015	0.314
Interpreting and critically appraising statistical information; n (%)	33 (86.8)	18 (75.0)	1.413	0.234
Using basic statistical techniques to analyse data; n (%)	32 (82.1)	15 (60.0)	3.798	0.051
Design and analysis of experiments; n (%)	26 (72.2)	13 (56.5)	1.544	0.214
Designing and conducting a questionnaire study; n (%)	20 (51.3)	11 (44.0)	0.323	0.570

As shown in Table 3, there was some evidence that PGRs (32/39, 82.1%) are more likely to attend a workshop on basic statistical techniques than staff members (15/25, 60.0%) (p=0.051). However, both groups appear just as likely to attend the other specified workshops as each other. This aligns with the findings presented in the earlier sections, suggesting that PGRs show more of a preference towards basic skills.

3.5. *Software preferences for quantitative research work*

To assess the demand for different statistical software packages (i.e. Excel, SPSS and R), all participants (52 PGRs and 36 staff members) were asked about their likelihood of using these for research work (Appendix A, Q11). Responses were grouped into likely (very likely or likely), not likely (very unlikely or unlikely) or unsure. Of the sample, 75 participants were likely to use Excel (85.3%), 51 were likely to use SPSS (58.0%) and 40 were likely to use R (45.5%) for their research work.

This was further explored across roles; PGRs or staff, as illustrated in Figure 4. For each software package, unsure participants were excluded due to the small sample sizes. Consequently, the total sample size for each software package varied across roles (i.e. Excel; 51 PGRs and 36 staff members, SPSS; 47 PGRs and 35 staff members and R; 43 PGRs and 34 staff members). Chi-squared tests were conducted to examine associations between role type and the likelihood of using each statistical software package (likely or not likely).

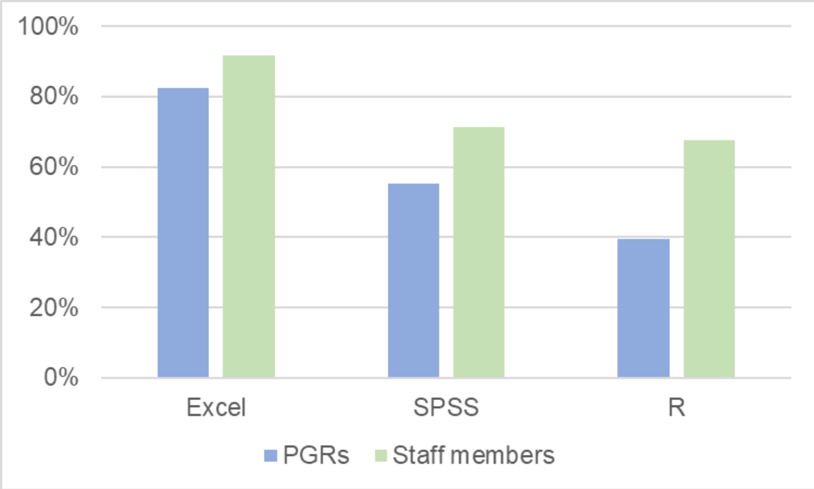

Figure 4: Participants who were likely to use each statistical software package by role; PGRs and Staff members

Excel usage appeared to be popular among both PGRs (42/51, 82.4%) and staff members (33/36, 91.7%) for research work, with no statistically significant association between role and likelihood of using Excel, $\chi^2(1, n=87) = 1.540$, p=0.215. For SPSS, staff members (25/35, 71.4%) were more likely to use it than PGRs (26/47, 55.3%), though this was not statistically significant, $\chi^2(1, n=82) = 2.214$, p=0.137. These findings may be due to the widespread familiarity with Excel and SPSS among both groups for research (and non-research) purposes.

In contrast, an association was found between the likelihood of using R and the type of role, $\chi^2(1, n=77) = 6.011$, p=0.014, with a greater proportion of staff members (23/34, 67.6%) indicating that they were likely to use R compared to the PGRs (17/43, 39.5%). Despite this, some staff expressed hesitation about using R for regular research work (Appendix A, Q15). One staff member commented, *"everyone seems to use R now and I find it intriguing but too complicated for occasional quant work. I'd rather use SPSS. However, an idiot's guide to R would be helpful!"* This suggests that while staff are aware of R, some may be reluctant to use this without further training or resources. Furthermore, one PGR requested workshops using R instead of SPSS, highlighting its relevance for their research (Appendix A, Q15). They commented, *"please use something like Python or R for the workshops. Some of the sessions look interesting but they are in SPSS which is really useless for me (and a lot of the PGRs in my centre)"*. Nevertheless, 11 respondents expressed reservations about using R (9 PGRs and 2 staff members), the most in comparison to the other software packages (i.e. Excel and SPSS).

3.6. Preferred mode of delivery (in-person/online) and months for attending research methods workshops

The review explored participants' preferences for workshop delivery modes (in-person or online) and the preferred months for attending such workshops (Appendix A, Q12 and Q13). Although accommodating everyone's preferences may be challenging, optimising the timing and format of workshops is crucial for encouraging attendance.

Of the sample (n=88), 55 participants preferred online workshops (62.5%), 20 suggested face-to-face (22.7%) and 13 were unsure (14.8%). When comparing these findings across the two roles; 65.4% of PGRs (34/52) and 58.3% of staff members (21/36) preferred online sessions, thus reflecting similar preferences; $\chi^2(2, n=88) = 0.456, p=0.796$. However, PGRs suggested options for *"watch[ing] recorded session"* and incorporating *"practical implementation… for better comprehension"* regardless of the delivery mode (Appendix A, Q15).

In addition, January and February emerged as the most popular months for training, with 47.7% of respondents favouring these months (42/88 respectively). However, nearly a third of participants were uncertain (27/88, 30.7%) about the best time for attending training. This was further explored based on role, as shown in Figure 5.

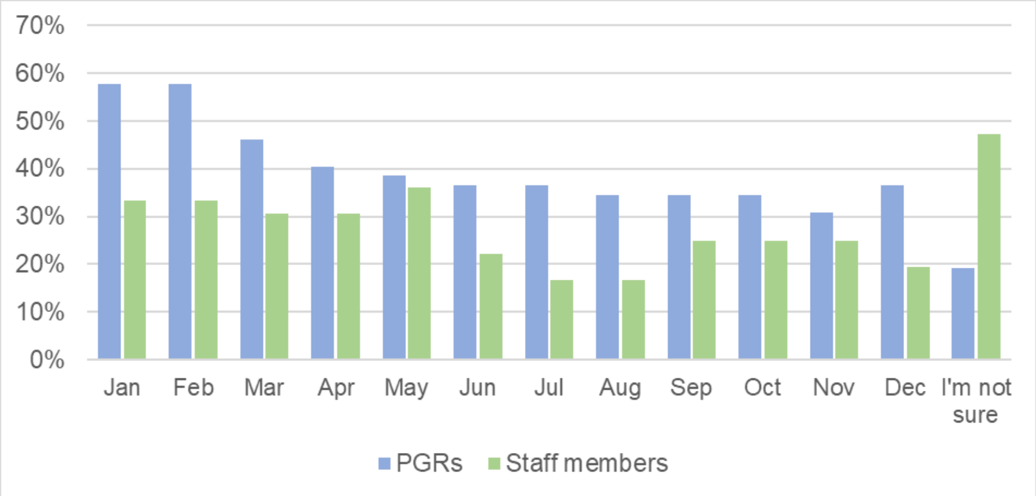

Figure 5: Participants preferred month for attending research methods workshops by type of role; PGRs (n=52) and Staff members (n=36)

Figure 5 illustrates that a higher proportion of staff members (47.2%) were uncertain about their preferred month for attending research methods training sessions compared to PGRs (19.2%). Using a chi-squared test, a statistically significant association was found between role type and uncertainty, $\chi^2(1, n=88) = 7.837, p=0.005$, likely due to work pressures and time constraints. Overall, staff seemed undecided with no clear favourite month, though they found June, July, August and December least favourable for attending training. This could possibly be due to annual leave during the summer and festive periods, as well as family commitments. Additionally, it appeared that staff members preferred *"agile"* and *"flexible"* offerings with *"advanced notice"* and *"more occurrences of each session"* (Appendix A, Q15). This difference is likely due to the nature of their roles, with staff constrained by various factors thus preferring more adaptable options.

For PGRs, any month appeared suitable, though they showed a preference for January and February. This may relate to the PGRs' start date, with nearly half beginning their programme of study in September (24/52, 46.2%). This timing likely reflects their need to assess and address training requirements a few months into their programme or following an annual progress review. Furthermore, one PGR highlighted the importance of training in the second year, suggesting January as an ideal time, *"the needs are urgent for many in year 2. If possible to timetable ASAP, such as in January would really help"* (Appendix A, Q15).

3.7. Supervisors' perceptions of PGRs training needs in quantitative skills

To gain a comprehensive understanding of PGRs' training needs, supervisors were surveyed about the skills and knowledge they feel their PGRs should develop in quantitative and qualitative research methods (Appendix A, Q14). Over two-thirds of the staff members with supervisory roles (66.7%, 24/36) provided insights, with 14 focusing on quantitative research methods.

Supervisors were keen for their PGRs to develop skills around statistical methods such as power calculations, mean/median calculations and understanding of parametric versus non-parametric tests, aligning with PGRs' responses in sections 3.2 and 3.3. For instance, one supervisor highlighted the need for PGRs to develop *"quantitative skills… simple statistics (means/medians etc), parametric/non-parametric analysis"*. Another supervisor outlined the value of developing skills in *"essential parametric and non-parametric methods – quantitative analysis. To Bland's Medical Statistics book level"*. Additionally, supervisors wanted PGRs to conduct *"basic statistics to interpret quantitative research"*.

However, differences were noted between supervisors' expectations and PGRs' perceptions of their needs. While supervisors wanted their PGRs to develop skills in statistical techniques, they also highlighted the importance of understanding experimental design, data organisation and coding, analysis interpretation and presentation of findings, areas not mentioned by the PGRs.

For example, one supervisor emphasised the need for skills in *"experimental design, power calculations, data organisation and coding for statistical analyses, interpretation of statistical tests [and] presentation of data"*. Another supervisor echoed these views, detailing essential skills for PGRs to develop. These included *"presentation of quantitative data in tables and text as appropriate in the discipline. What to consider when designing a study such as a survey (e.g., sampling technique, questionnaire design). Preparation of a data analysis plan. How to document statistical analysis done, including code used to analyse the data…."*.

Moreover, supervisors indicated that training needs can be *"project specific"* with one supervisor noting the difficulty of defining training needs in advance. They stated, *"it is very difficult to define training needs beforehand as I always think it needs to be aligning with a project you are working on…"*.

Supervisors also highlighted the importance of timely training and the drawbacks of not applying learning when needed. They commented, *"…some basic knowledge is needed to decide upon the best research method but actual training needs to happen shortly before or during data collection more towards analysis phase. Research method skills need to be maintained or used frequently otherwise it will sink to the back of my mind and I won't know how to apply it by the time it is relevant and needed"*.

Additionally, supervisors expressed the value of familiarising PGRs with *"software packages"* and *"quantitative terminologies"*. They felt that having this knowledge supports good practices in statistical methods and helps *"correct common bad habits"*.

4. Discussion and Summary: The Coventry Perspective

Reflecting on the responses outlined by participants in this review and comparing them with what we currently do in **sigma**, Maths and Stats Support at Coventry University, has been insightful. The findings provide a clearer direction for **sigma**'s delivery and identify the statistical skills and techniques that our PGRs and staff feel they need to develop. This section outlines key findings and how they align with **sigma**'s current offering, as well as changes already implemented to support researchers' development needs. We hope these findings can aid other institutions and support practitioners in planning content and identifying quantitative training for their PGRs and staff members.

4.1. Reflections on the findings

The findings provide some reassurance about our current offering in **sigma**, with 59% of respondents expressing a need to develop basic skills. However, the results suggest that there is also some demand for training courses in more advanced statistical techniques, particularly among staff members. This could, of course, be due to those who responded to the survey having particular experience with statistical methods in their own research, but it appears we should consider tailoring programmes, perhaps targeting introductory workshops more towards early stage PGRs.

Currently, in **sigma**, we offer workshops on basic skills and introductory statistical methods, including one-way ANOVA and simple linear regression. We also direct PGRs and staff to internal and external resources for advanced methodologies. For example, the **sigma** website (https://libguides.coventry.ac.uk/sigma/statsresources) features resources on two-way ANOVA, panel data regression and meta-analysis. However, the findings suggest a potential demand for us to expand our workshop offering to cover more advanced statistical techniques, such as logistic regression, Bayesian statistics and structural equation modelling. We are currently looking to develop self-study resources on these topics.

Beyond basic training in how to do statistics, PGRs, in particular, expressed an interest in developing skills to understand and critically appraise quantitative evidence. This fits with our experience in **sigma** where we have seen a rise in the number of students conducting meta-analysis and systematic review-type projects since the Covid-19 pandemic. Our current training in this area is limited, suggesting there is scope to develop resources and provide training in this skill, especially as it cuts across many research disciplines.

The findings also indicated that both staff and PGRs were keen to enhance their understanding of essential statistical techniques for research projects, including interpreting p-values and addressing violations to statistical assumptions, as well as issues relating to study design such as performing power calculations and calculating effect sizes. At present, we offer workshops which incorporate explanations of p-values and how to handle violations of statistical assumptions. However, there is an opportunity to expand our offering in relation to performing power calculations and calculating effect sizes, which is not something we currently focus on.

In addition to developing skills in a range of statistical techniques, supervisors highlighted other important skills for PGRs to develop when undertaking research projects. These included data organisation and preparation, including coding of variables, as well as interpreting findings and presenting them appropriately. Therefore, it may be beneficial to consider resource development in these areas, with links to resources around study design, as these skills are applicable across a range of disciplines.

When investigating opinions around software packages, participants had a mixed response to using R for research. This may be due to various factors, such as the programming skills required, which can be challenging for those unfamiliar with coding. Early career researchers, in particular those without a maths and stats background, may be unfamiliar with R and could find learning R a steep learning curve, explaining why fewer PGRs indicated its use. Staff members seem more likely to use R, potentially due to their research experience and familiarity with a wider range of software packages.

From our experience at Coventry University, many courses and programmes are increasingly moving towards using R as a software package for research work, particularly in social science disciplines. This shift is likely due to R's adaptability, flexibility and capability to handle advanced statistical techniques and large data sets, making it a preferred software for statistical data analysis (Li, 2018). Additionally, R is free and open-source, eliminating the costs associated with purchasing and renewing licenses; as such, this is a cost-effective option for students, staff and the institution (SAGE Campus, 2019). Here in **sigma**, we currently offer an introductory workshop on R. However, our findings suggest that additional R workshops could be considered since usage of this software is increasing, particularly among staff members.

Furthermore, in **sigma**, we do not currently offer workshops on Excel as most people already have some level of familiarity with this software. Instead, support is provided through drop-ins or one-to-one appointments, and it tends to be at a basic level, not moving beyond producing tables and charts. Given the popularity of Excel, it may be worthwhile to review our support relating to this package as a go-to for basic analyses.

*4.2. Changes made in **sigma**, Maths and Stats Support Centre*

Resource and content planning is an ongoing process and we will continue to make adaptations to improve our statistics workshops and training resources. However, in response to the survey, we have already implemented a few small changes as shown in Table 4. For instance, we have removed workshops on study design, which was quite generic and did not incorporate elements such as power calculations, and questionnaire design due to low demand. We have merged some existing workshops and introduced a new workshop on choosing the right statistical test to help researchers plan and explore potential statistical methods, since this is a topic that we, at least anecdotally, have seen demand for. We have updated the titles and descriptions of the workshops to help attendees make more informed decisions about the suitability of the sessions.

Additionally, recognising that online delivery for research methods workshops is preferred by both PGRs and staff, we have transitioned to delivering all workshops online. This came with challenges, especially for software-related content, though we reviewed and adapted the material, ensuring online delivery was suitable. For example, if the sessions make use of a software package, we have incorporated interactive demonstrations and have included some time at the end for participants to have a go at using the software themselves. This approach encourages engagement while providing a meaningful learning experience. We also encourage attendees to obtain in-person support through our drop-in sessions or one-to-one appointments when needed.

Additionally, we have rescheduled the workshops to late January and February and start promoting them before Christmas to help increase attendance. We plan to repeat the workshops in May to maximise training opportunities for both staff and PGRs across the year.

With these small adjustments and future resource development (e.g. topics around understanding published results, different regression techniques, workshops using R, etc), we aim to provide PGRs and staff with the essential quantitative skills required for undertaking research. The survey gave insights into the differing needs of PGRs and staff and we will take this into consideration when designing and promoting future workshops, with foundation-level workshops perhaps aimed more at the PGR group.

Table 4: **sigma** workshop offering before and after the review

Before Review – 9 workshops	**After Review – 7 workshops**
Study Design and Statistical Terminology	Workshop removed
Introduction to Questionnaire Design	Workshop removed
Descriptive Statistics	Workshop title changed to: Understanding Descriptive Statistics
Introduction to SPSS26	Workshop title changed to: Getting Started with SPSS
Introduction to R for Windows (Using RStudio)	Workshop title changed to: Getting Started with R and RStudio
Introduction to Statistical Inference	Workshop title changed to: Understanding Statistical Inference – What is a p-value?
Introduction to Analysis of Variance (ANOVA)	Workshop title changed to: Comparing Groups
Introduction to Non-Parametric Statistics	Workshop removed
Correlation and Regression	Workshop title changed to: Finding Relationships
-	New workshop added: Choosing the Right Test

5. Appendices

5.1. Appendix A – Survey Questions

1. Which of the following best describes you?

Postgraduate Researcher (PGR)
Staff member who is also a Postgraduate Researcher (PGR)
Staff member
Other

1.a Please provide details of your course of study e.g. the topic area, brief details of your research (free text).

1.b Are you a full-time or part-time PGR?

Full-time
Part-time

1.c Please state when you commenced your programme of study (free text).

1.d Please state the expected end date of your programme of study (free text).

1.e What year of study are you in?

Year 1	Year 2	Year 3	Year 4
Year 5	Year 6	Year 7	Other

1.e.i If you selected Other, please specify (free text):

2. Do you belong to a Research Centre?

Yes
No
Not sure

2.a Which Research Centre do you belong to?

Centre for Agroecology, Water Resilience	Centre for Arts, Memory and Communities
Centre for Business in Society	Centre for Computational Science and Mathematical Modelling
Centre for Dance Research	Centre for E-Mobility and Clean Growth
Centre for Financial and Corporate Integrity	Centre for Fluid and Complex Systems
Centre for Future Transport and Cities	Centre for Global Learning
Centre for Healthcare Research	Centre for Intelligent Healthcare
Centre for Manufacturing and Materials	Centre for Postdigital Cultures
Centre for Sport, Exercise and Life Sciences	Centre for Trust, Peace and Social Relations
Other	

2.a.i If you selected Other, please specify (free text):

2.b Please state the Faculty/School/Area of the University you are located in (free text):

3. For each of the following research methodologies, how much knowledge do you feel you need to have in order to carry out your own research/work? Please select one response per row. Measured on a 5-point scale: No Knowledge, Basic Knowledge, Good Working Knowledge, Advanced Knowledge, I'm not sure.

Quantitative research (e.g. working with quantitative data such as survey responses, understanding and interpreting statistical information, data from a planned experiment etc.)
Qualitative research (e.g. working with textual/descriptive data from interviews, observation, documents, focus groups etc.)
Mixed methods research (i.e. a combination of both quantitative and qualitative research)

4. The next few questions relate to building quantitative skills, such as working with and/or interpreting statistical information. If you are certain that you do not need to develop quantitative skills in your research/work (e.g. if you are a purely qualitative researcher or you already have the required quantitative skills), you will be able to skip these questions. Please select the correct option below.

I do not need to develop quantitative skills in my research/work
I may need to develop some quantitative skills

5. In relation to quantitative methods, which of the following skills do you think you might need to learn? (Select as many as apply to you):

Understand statistical outputs reported in publications, reports, books etc.
Replicate quantitative work that others have done
Appropriately design my research
Understand methods for collecting quantitative data
Apply basic statistical techniques
Apply advanced statistical techniques
I'm not completely sure what my skills requirements are but I'm likely to need some skills in quantitative methods
Other

5.a If you selected Other, please specify (free text):

6. Do you have any idea of specific statistical techniques you need to know about and/or use in your research? Please list as many as you can think of or simply state, "I don't know". For example, confidence intervals, t tests, ANOVA, descriptive statistics, regression analysis etc (free text).

7. If workshops were offered to you in the following skills, how likely would you be to attend? Please select one response per row. Measured on a 5-point scale: Very Unlikely, Unlikely, Quite Likely, Very Likely, Not sure.

Interpreting and critically appraising statistical information
Designing and conducting a questionnaire study
Using basic statistical techniques to analyse data
Using advanced statistical techniques to analyse data
Using statistical software
Design and analysis of experiments

8. The next two questions relate to building qualitative skills, such as working with textual data from interviews, focus groups, ethnographic research etc. If you are certain that you do not need to develop qualitative skills in your research/work (e.g. if you are a purely quantitative researcher or you already have the required qualitative skills), you will be able to skip these questions. Please select the correct option below.

I do not need to develop qualitative skills in my research/work
I may need to develop some qualitative skills

9. In relation to qualitative methods, which of the following skills do you think you might need to learn? (Select as many as apply to you):

Appropriately design my research
Develop an appropriate theoretical framework for my research
Use appropriate data collection methods
Apply appropriate data analysis techniques
I’m not completely sure what my skills requirements are but I’m likely to need some skills in qualitative methods
Other

9.a If you selected Other, please specify (free text):

10. Do you have any idea of specific qualitative methods/approaches/techniques you need to know in your research? Please list as many as you can think of or simply state, “I don’t know”. For example, grounded theory, content analysis, narrative analysis, discourse analysis, ethnography, phenomenology etc (free text).

11. How likely are you to use the following software packages in your research/work? Please choose one response per row. Measured on a 5-point scale: Very Unlikely, Unlikely, Quite Likely, Very Likely, Not sure

SPSS
R
Excel
NVivo

11.a Other software packages, please specify (free text):

12. What format do you prefer when attending workshops?

In-person (Coventry University campus)
Online
Not sure

13. When would you be most likely to attend research methods workshops such as those mentioned in this survey? (Select as many as apply to you):

Jan	Feb	Mar
Apr	May	Jun
Jul	Aug	Sep
Oct	Nov	Dec
I’m not sure		

14. Do you have supervision responsibilities for any students undertaking postgraduate research?

Yes
No

14.a Please outline the knowledge and skills related to quantitative and/or qualitative research methods that you would like your Postgraduate Researcher to develop/have? If you're not sure, please state this (free text).

15. Is there anything else you would like to share about training requirements for your research/work? Please comment below (free text).

5.2. *Appendix B: Statistical techniques participants felt they might need to know about and/or use for their research work mentioned infrequently (n=66)*

Statistical technique	**Frequency**
Standard deviations and errors	1
Wilcoxon test	1
Log-linear analysis	1
Maximum likelihood	1
Covariate variables	1
Dummy coding	1
Experimental analysis	1
Exploratory analysis	1
Geographical information systems	1
Modelling	1
Network approach	1
Path analysis	1
Pre and post-hoc power analysis	1
Probability distribution	1
Gaussian distribution	1
Interclass correlation coefficient	1
Meta-analysis and funnel plots	1

6. References

Braun, V. and Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, 3(2), pp.77–101.

British Academy (2012). *Society counts: Quantitative skills in the social sciences and humanities*. [online] Available at: https://www.thebritishacademy.ac.uk/documents/206/Society-Counts-Quantitative-Skills-in-the-Social-Sciences-Humanities-Position-_3xJi9mM.pdf [Accessed 15 Sep. 2024].

British Academy (2015). *Count us in: Quantitative skills for a new generation*. [online] Available at: https://www.thebritishacademy.ac.uk/documents/220/Count-Us-In.pdf [Accessed 14 Sep. 2024].

ESRC (2022). *ESRC Postgraduate Training and Development Guidelines Third Edition (2022)*. [online] Available at: https://www.ukri.org/wp-content/uploads/2015/09/ESRC-281123-PostgraduateTrainingDevelopmentGuidelines2022.pdf [Accessed 14 Sep. 2024].

Field, A.P. (2018). *Discovering statistics using IBM SPSS statistics*. 5th ed. Los Angeles: Sage.

Krippendorff, K. (2018). *Content Analysis: an Introduction to Its Methodology*. Sage Publications.

Lawson, D., Grove, M. and Croft, T. (2019). The evolution of mathematics support: a literature review. *International Journal of Mathematical Education in Science and Technology*, 51(8), pp.1224–1254. https://doi.org/10.1080/0020739x.2019.1662120.

Li, Q. (2018). *Using R for Data Analysis in Social Sciences*. Oxford University Press.

SAGE Campus. (2019). *Why universities are switching to R for teaching social science*. [online] Available at: https://campus.sagepub.com/blog/why-universities-are-switching-to-r-for-social-science [Accessed 21 Aug. 2024].

Vitae (2011). *Vitae Researcher Development Framework 2011*. [online] Available at: https://www.vitae.ac.uk/vitae-publications/rdf-related/researcher-development-framework-rdf-vitae.pdf/view [Accessed 14 Sept. 2024].

THIS PAGE DELIBERATELY LEFT BLANK

RESEARCH ARTICLE

Enhancing Statistics Support with Artificial Intelligence

Ben Derrick, Mathematics and Statistics Research Group, University of the West of England, Bristol, UK. Ben.Derrick@uwe.ac.uk

Iain Weir, Department of Computer Science and Creative Technologies, University of the West of England, Bristol, UK. Iain.Weir@uwe.ac.uk

Abstract

The integration of artificial intelligence (AI) technologies is revolutionising traditional methods of teaching and learning. The University of the West of England, Bristol, has developed a generative AI policy that encourages AI literacy, personal learning and creativity. In accordance with this policy, we demonstrate use of AI within an established help drop-in service at the university. Data analysis advice from statisticians is provided to students via a newly formed 'Stats Clinic' which aims to act as a triage service within the institution's existing 'espressoMaths' service, open to all.

With appropriate student preliminary engagement, including the use of AI, the productivity and value of student-academic discussions can be greatly increased. Detail is given of how students can use artificial intelligence to get the most out of pre-visit engagement and therefore ultimately their visit with a statistics professional.

Examples where students have applied varying levels of engagement with pre-visit recommended actions are discussed, with empirical evidence from the sessions indicating that those embracing AI are more aware of their data analysis and can comprehend advice more readily.

Keywords: drop-by station, statistical test selection, ChatGPT, generative AI.

1. Introduction

In today's educational landscape, the integration of artificial intelligence (AI) technologies is revolutionising traditional methods of learning. Educational institutions and academics are increasingly encouraged to adopt AI-driven solutions to enhance student engagement, improve learning outcomes, and streamline administrative processes (Hargrave, Fisher and Frey, 2024).

The authors of this paper have been involved in the discussions relating to the principles for using generative artificial intelligence, as part of the AI Community of Practice and Working Group at the University of the West of England, Bristol (UWE Bristol). These principles were formally adopted during the 2023-2024 academic year. "*UWE Bristol is committed to harnessing the transformative potential of generative AI to enhance learning, teaching and assessment. We aim to support students and staff to become AI-literate, equipped to drive progress and innovation through the ethical use of these powerful technologies.*" (UWE Bristol Principles for using generative artificial intelligence, 2024).

This paper focuses on the development and implementation of a drop-by station, ‘Stats Clinic’, specifically designed for handling statistics queries in a university setting, supported by preliminary student engagement with ChatGPT (OpenAI, 2023). The aim of this initiative is to provide students with an efficient and accessible platform to seek assistance with their statistical inquiries before engaging with human tutors or faculty members. By leveraging AI technologies, students can receive immediate feedback, access resources, and gain insights into complex statistical concepts, thereby facilitating a deeper understanding of the subject matter.

2. Previous offering and rationale for change

Since 2008, the university has offered a mathematics and statistics drop-in session called espressoMaths. The service is open to all members of the university community and is usually held around lunchtime each weekday during term time. The physical presence is held in various heart zones of the university, is accessible and user-friendly. It allows spontaneous attendance, no booking system is in place. As such it is designed that interactions are brief to ensure that no student must wait more than 5-10 minutes. The website provides detailed schedules where mathematicians or statisticians are available so that users can visit the most appropriate member of the espressoMaths team.

The most common query Statisticians are asked at espressoMaths and via other communications such as email is how to analyse a dataset. Often users:

- struggle to articulate their research question and study design;
- have had little education in statistics or have not fully engaged with the teaching;
- expect to be performing ‘advanced’ statistical techniques within minutes;
- are passive users of statistics expecting detailed instructions from the academic.

There was a need to develop a consistent and efficient approach to assisting users of statistics. Some statistics advisors on the espressoMaths service have previously spent significant time guiding users through an entire process in a way that restricts student engagement and is poor use of the time. Other providers of the service consider themselves to be part of a triage service identifying user needs and prescribing solutions, akin to the statistician being the ‘General Practitioner’ and the visitor being the ‘patient’. However, this has been a challenge for patients with extensive needs and there are many patients requiring repeat prescriptions. Records show that over half of all ‘espressoMaths’ interactions are 20 minutes or more, with many lasting a full hour; this conflicts with the founder’s original ‘espresso’ vision of timely assistance (Henderson and Swift, 2011).

Modern advances in technology and the challenges faced in dealing with statistical queries mean that the existing structure of espressoMaths is no longer adequate for handling data analysis queries. Given the established brand of espressoMaths and its reputation for adding value throughout the university, the development of a Stats Clinic is best initiated and grown within this framework. Additionally, statistics academics receive numerous ad-hoc queries via email that lack a recognised management process. A Stats Clinic could provide a structured outlet for these inquiries.

Some preliminary work by the user would help manage the high demand on statisticians and reduce the need for extensive initial effort. This preparation encourages users to clearly articulate their research questions and study design. The Stats Clinic aims to oversee and support students in using ChatGPT to identify appropriate analyses for their data.

It is important to note that we are not suggesting that ChatGPT can be used to do work for the student, instead we suggest it is used to gather advice. Likewise, we are not proposing that ChatGPT can replace a qualified statistician, instead it can be used to assist in the teaching process.

3. Stats Clinic offering

Following consultation with the espressoMaths coordinator and the statistics team, the Stats Clinic at UWE Bristol was established and has been running since October 2023. A summary of the process is given in Figure 1.

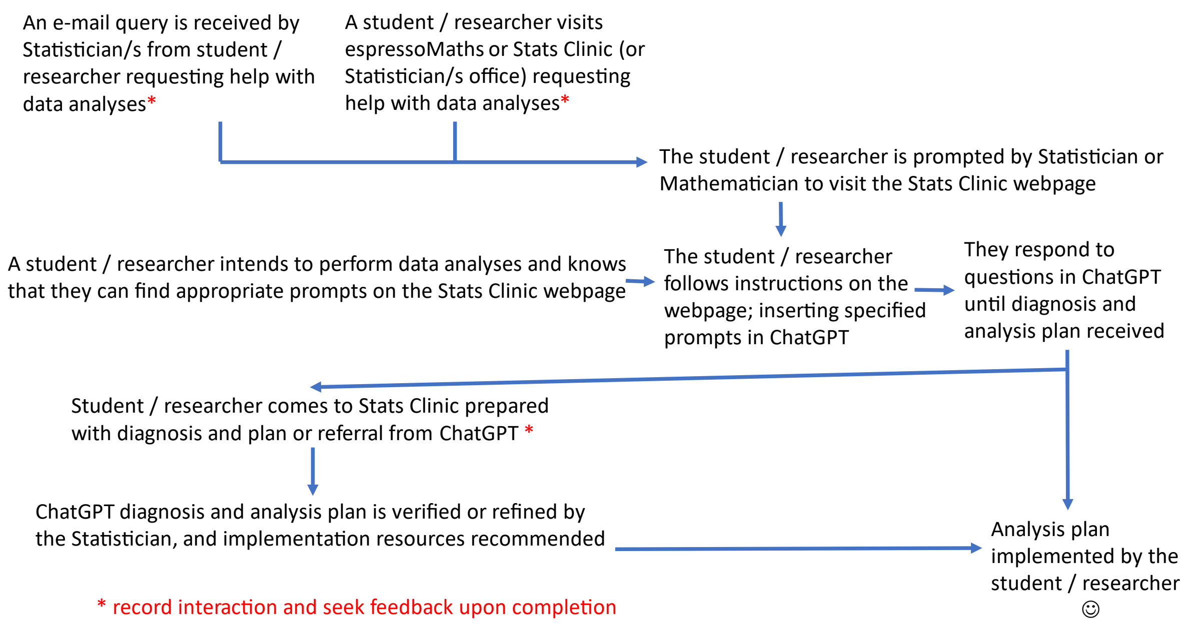

Figure 1. Stages of engagement in the Stats Clinic process.

We aim to introduce some concepts of ethical use of AI by describing an appropriate use of AI as a learning tool. Given in Figure 2 is an extract of the instructions to those with data analyses related queries, as provided to users on the espressoMaths website (UWE Bristol Mathematics and Statistics Study Skills, 2024). This details the preliminary engagement required prior to attending a Stats Clinic session. The prompts are an extension of those presented by Owen (2023), and are similar to the prompts developed by Goodale (2024).

Students are directed to seek further guidance relating to responsible use of AI from their module handbook or assessment guidelines.

1. **Understanding your research question**

Prior to your interaction with ChatGPT (or a Stats Clinic advisor) you will need to identify your research question. You should have an awareness of what you are trying to measure or predict (response variable), and what factors you have or will collect data for that could impact measurements of that variable (explanatory variables). When you have an understanding of what you are trying to achieve, then you will be in a position to interact with ChatGPT.

Recommended prompts

Insert the following prompt into ChatGPT:

I am going to ask a question about statistics. When answering me use the following approach: 1. Make your explanations comprehendible by an undergraduate degree student 2. Ensure your responses are precise and based on recognized knowledge. 3. Consult a variety of sources and contexts. 4. Avoid advancing societal stereotypes or biases. 5. State uncertainty if the answer isn't clear-cut. 6. Focus strictly on the subject without digressing. 7. Recommend exploratory data analyses and statistical tests I should perform. 8. Ask me questions one at a time until you thoroughly understand my research question and study design before providing a solution.

Here is my research question: **[insert details about what you are trying to achieve and what data you have / will have]**

ChatGPT will ask some important questions to gain an understanding of your research question and data collected. Please respond to each question fully, asking ChatGPT for clarification of a question where required.

ChatGPT will provide a solution, read and review its recommendations. We suggest that you further interact with ChatGPT using some of the following prompts that we have found useful:

- Please show me how to do **[insert name of analysis recommended by ChatGPT]** in **[insert the name of the statistical software you use]**;
- Please inform me how to interpret the results of **[insert name of analysis recommended by ChatGPT]**;
- What do you mean by **[insert any terminology used by ChatGPT that you do not understand]**?;
- Is there anything else that I should consider?

If you are happy with the solution then there may be no need to visit the statistician, however you may still wish to do so to check the advice given.

You may receive a referral from ChatGPT to see a statistics specialist, or would otherwise like further verification of the response from ChatGPT from a statistician. If this applies then please do visit the Stats Clinic in person. When visiting the Stats Clinic you should bring your ChatGPT conversation with you to speed up the process in the Stats Clinic. Please be advised that if you visit the Stats Clinic without evidencing your interaction with ChatGPT, we will direct you to this page to complete the suggested prompts in ChatGPT.

Figure 2. Stats Clinic preliminary engagement instructions on espressoMaths website.

4. Typical interactions during drop-in sessions

Users at Stats Clinic were generally undergraduates studying non-statistical subjects, seeking assistance with a final-year dissertation or a second-year statistics analysis project.

The following summarises the typical types of user who present at the Stats Clinic and the effectiveness of the service for each type.

1. The Traditionalist

Occasionally, students and staff revert to traditional methods used before the establishment of the Stats Clinic, such as discussing problems step-by-step and recommending books. These interactions often lead to students attending multiple sessions, relying on traditional textbooks for reference, and engaging in time-consuming exchanges. The repeated need to cover the same concepts suggests that the communication may be ineffective.

2. The Chancer

Many users had not initially engaged with preliminary activities, including those seeking step-by-step guidance. Statisticians directed these students to complete preliminary activities and use ChatGPT, which was generally well received. This approach effectively encourages students to become self-sufficient by utilising AI tools.

3. The Premium Payee

A few students engaged with AI before visiting the Stats Clinic by subscribing to premium AI services not recommended by us. These students reported that the paid version of ChatGPT (Century) provided inconsistent results and caused further confusion. This indicates that our preferred, widely accessible approach is at least equally effective, and there appears to be no clear advantage currently for those who opt for more expensive services. However, user experiences with AI may vary, including the value of paid versions. The relative benefit of different versions of AI and the potential disparity between those that can afford premium versions and those that cannot, is something that should be continually monitored.

4. The AI Apprehensive

Some users did not engage with preliminary activities due to concerns about AI accuracy. During interactions, guiding these users through data analysis with ChatGPT and verifying the results led to positive feedback about AI's usefulness. Demonstrating AI's reliability can ease apprehensions and encourage greater engagement.

5. The Casual AI User

Some students were eager to interact with ChatGPT but do not use it effectively, especially those who do not follow suggested prompts and end up confused by the results. This confusion is often resolved by encouraging students to think more carefully about their ChatGPT prompts. Ensuring that students use appropriate prompts and understand context-specific nuances is essential for effective AI use.

6. The Fully Engaged

Several users engaged deeply with preliminary activities and followed the advice as intended. These users demonstrated a high level of comprehension and used the service to confirm their understanding. This straightforward process also allowed time for discussing concepts beyond the basics, guided by AI. This highlights the potential of AI tools to enhance learning outcomes when fully utilised, and the additional value a statistician can provide once the basics are covered by AI.

5. Discussion

The number of users visiting the Stats Clinic subsidiary of espressoMaths in the 2023-24 academic year is estimated to be around 30-40. We believe this represents a drop in the frequency of visits based on previous iterations of the statistics provision at espressoMaths. The number of users being helped by the AI prompts we provide without then visiting a statistician in-person is difficult to quantify, it is likely that many users are benefiting from the service without direct interaction.

Initial concern that the student body might view the approach as ‘lazy’ or ‘cost-cutting’ has been alleviated by the highly supportive student reaction to the service, verbal student comments during the sessions include:

“It’s great to see someone using it [ChatGPT]”;

“Very useful for my project, great use of ChatGPT showing me how to use it correctly and efficiently to help me solve a range of statistical problems and data analysis problems”;

“ChatGPT should be used in all lectures, Education is always behind on new technology”;

“I wouldn’t have been able to do this without ChatGPT”;

“That’s so cool! ChatGPT is useful for some things. Fab. I was going to email but this is really useful”.

Formal written feedback is requested on the website but is only recorded if the user chooses to engage with the survey. The number of users providing formal written feedback has been low, but overwhelmingly positive as shown in Table 1.

Table 1. Number of responses to the question, “How would you rate your espressoMaths experience today?”, in each of the last three academic years.

Academic year	Extremely positive	Somewhat positive	Neutral	Somewhat negative	Extremely negative
2023-2024	12	2	0	0	1
2022-2023	15	0	0	0	0
2021-2022	22	4	0	0	0

The student who responded negatively stated:

“*I would like it if the session was more interactive and helpful […] he seemed like he didn't want to help me much”.*

The negative feedback suggests that there may be a stigma to overcome for some students regarding the use of AI, and the perceived reduction in human contact.

The stigma against AI is not confined to some sections of the student body, many staff are also yet to fully embrace AI capabilities. In one interaction, a PhD student arrived with what the statistician viewed to be important research; given the level and importance of the research the statistician *'in the moment felt too embarrassed to suggest ChatGPT'*, even though all the questions and approaches were straightforward and could have been handled by AI more efficiently than the statistician. Many academics are not mentioning AI at all. Further research into staff perceptions of AI, and development of staff training strategies, is required.

Our experience is that when students use our suggested prompts, the responses from ChatGPT have excellent accuracy. However, limitations of the approach are that we do not know how many students are using AI as advised, and we cannot be sure of the accuracy of the guidance received by those that do not seek verification from an expert statistician.

6. Conclusion and Recommendations

Most of the demand for the Stats Clinic drop-in sessions is related to final year projects. Project-Based Learning (PBL) plays a crucial role in modern education by offering students immersive, hands-on experiences that go beyond rote memorisation and encourage critical thinking, problem-solving, and independent thought (Derrick and Weir, 2024). Integrating AI-driven drop-in sessions within PBL frameworks can significantly enhance the learning process.

AI provides real-time support and guidance to students as they navigate complex project challenges, offering insights, resources, and feedback tailored to individual needs. By leveraging AI capabilities such as automated advice and feedback, drop-in sessions can empower students to delve deeper into project exploration, refine their ideas, and develop essential skills in a supportive and interactive environment. This integration not only enriches the PBL experience but also prepares students for future roles where AI technologies are increasingly prevalent, fostering a holistic and adaptive approach to learning.

The Stats Clinic case study at UWE Bristol demonstrates the transformative potential of AI in enhancing academic support services and improving student outcomes. By leveraging AI technologies responsibly and ethically, educational institutions can create inclusive learning environments that empower students and foster their success. As AI continues to evolve, maintaining a balance between technological innovation and ethical considerations is paramount to ensure AI remains a catalyst for positive change in education. The Stats Clinic approach and similar initiatives should evolve over time as the AI landscape evolves, including new examples and dynamic guidance.

Based on the experiences and insights gained from the Stats Clinic, several recommendations and future directions emerge:

- Promoting AI literacy among students and faculty to maximise the benefits of AI integration, including research into staff and student perceptions;
- Conducting ongoing research and development in AI-driven educational technologies to address evolving student needs;

- Fostering partnerships and collaborations within the AI and education ecosystem to promote responsible AI adoption and innovation. This includes advancing the role of the Stats Clinic as a hub for further education and development in these areas.

Students that we have engaged with throughout this transition generally seem to support the use of AI assistance. Those who follow the Stats Clinic instructions and fully embrace the AI technology have had the most successful interactions and are better prepared for future work environments. In conclusion, incorporation of AI has allowed us to adapt to evolving student needs, provide real-time feedback that helps students learn at their own pace, and reduce statistician time required resulting in staffing cost efficiencies.

7. References

Derrick, B., & Weir, I. 2024. Project-based learning integrated with e-Assessment. MSOR Connections, 22(2), 12-20. https://doi.org/10.21100/msor.v22i2.

Goodale, T. 2024. Using Generative AI to help with statistical test selection and analysis. MSOR Connections, 22(3). https://doi.org/10.21100/msor.v22i3.1485

Hargrave, M., Fisher, D., & Frey, N. 2024. The Artificial Intelligence Playbook: Time-Saving Tools for Teachers that Make Learning More Engaging. Corwin Press.

Henderson, K., & Swift, T. 2011. espressoMaths: A drop-by station. MSOR Connections, 11(2), 10-13)

Owen, A. 2023 Statistics Support Using ChatGPT: the good, the bad and the.... CETL-MOSR Conference, Cardiff University, 2023.

OpenAI. 2023. ChatGPT (GPT-4) [Large language model]. https://chat.openai.com/chat [accessed 1/09/2023]

UWE Bristol Mathematics and Statistics Study Skills. Available from: https://uwe.ac.uk/study/study-support/study-skills/mathematics-and-statistics [accessed 9/07/2024]

UWE Bristol Principles for using generative artificial intelligence. Available from: Principles for using generative artificial intelligence (AI) - Academic information | UWE Bristol [accessed 12/08/2024]

RESEARCH ARTICLE

Exploring the use of AI in mathematics and statistics assessments

Siri Chongchitnan, Warwick Mathematics Institute, University of Warwick, Coventry, UK.
Email: Siri.Chongchitnan@warwick.ac.uk
Martyn Parker, Department of Statistics, University of Warwick, Coventry, UK.
Email: Martyn.Parker@warwick.ac.uk
Mani Mahal, Department of Statistics, University of Warwick, Coventry, UK.
Email: Mani.Mahal@warwick.ac.uk
Sam Petrie, Warwick Mathematics Institute, University of Warwick, Coventry, UK.
Email: Sam.Petrie@warwick.ac.uk

Abstract

The mathematical sciences and operational research (MSOR) community in higher education is still largely unprepared to adapt to the rapid rise of generative artificial intelligence (genAI) and its impact on assessment strategies. Whilst in-person exams remain an essential assessment mode for MSOR, take-home assignments are also an integral assessment tool. This work investigates concerns that current assignments are not robust against genAI and the way students use genAI. In this work, we address the following questions: 1) How well can genAI perform in current assignments? 2) To what extent do students currently use AI in take-home assignments? 3) How should assessment strategies evolve given the rapid improvement of genAI? Our research involves an investigation of genAI's performance in a range of MSOR assignments. We also conducted surveys and discussions with mathematics and statistics students and staff at the University of Warwick. We make recommendation and conclude that genAI represents a catalyst for innovation and assignments, perhaps adapted, should remain a core assessment in MSOR.

Keywords: Generative Artificial Intelligence, Mathematics, Statistics, Assessments.

1. Introduction

The mathematical sciences and operational research (MSOR) community, like all disciplines in higher education, needs to address the rapid integration of advanced AI technologies into academic environments. While in-person examinations have traditionally been the primary method of assessment in these disciplines, take-home assignments remain a critical component for evaluating student knowledge and problem-solving skills (Iannone and Simpson, 2011, 2012, 2022). The emergence of genAI presents challenges to the integrity of these assignments.

The primary purpose of this work is to explore and understand the impact of generative artificial intelligence (genAI) on mathematical assessment focusing on the (paid) large language model (LLM) GPT-4o. By examining the capabilities and limitations of GPT-4o, this article project aims to provide insights that will inform assessment strategies within the MSOR community. Initial evaluation of other genAI models demonstrated that GPT-4o provided the best responses, thus this work focuses on this model.

The research presented in this work was carried out at the University of Warwick, a large UK university where there are around 2000 taught (Undergraduate (UG) and postgraduate (PGT)) students in the mathematics and statistics departments.

This paper covers three areas.

1. How is AI's performance on current assignments? Evaluate GPT-4o's ability to solve university-level mathematics and statistics assignments.

2. Examine students' use of AI. Determine how are students using genAI to complete their assignments. What are their perceptions and understandings of these tools?

3. Assessment strategies. Discuss how assessment strategies should evolve, given the rapid improvement of genAI.

The full report for this work is available online (Chongchitnan, et al., 2024).

2. The Emergence of ChatGPT and Its Impact

ChatGPT (OpenAI, 2022) fundamentally altered the educational landscape virtually overnight. Students could suddenly, instantly, and for free, obtain answers that far exceeded the capabilities of AI task managers or search tools like Siri or Google Assistant. This shift raised concerns about the integrity of academic assessments, particularly in essay-based subjects where students could easily generate large portions - or even entire assignments - within seconds.

At the time of its release, ChatGPT was powered by a single LLM: GPT-3.5. This model quickly became synonymous with the ChatGPT brand and remains, according to our study, the most popular version used by students nearly two years later, despite being replaced by GPT-4o mini. GPT-3.5, like GPT-4o mini, was always offered for free with usage limits.

GPT-3.5 capabilities are limited by its training data, which often includes both accurate and inaccurate information (OpenAI, 2022; Huang et al., 2023). This limitation affects its performance in mathematical contexts, where rigorous logic and structured reasoning are required through multiple steps.

Since LLMs generate answers to mathematical problems through the same probabilistic mechanism used for text generation, it is not unusual to find counting or other basic mathematical errors. OpenAI provided a generic warning at the bottom of all chats that "*ChatGPT can make mistakes*". This phenomenon, where the model produces responses that seem accurate or correct but are underpinned by flawed reasoning, is known as *hallucination.* As a result, many students who initially experimented with GPT-3.5 developed a negative perception of the capabilities of LLMs broadly, but particularly in MSOR subjects (Attewell, 2024; Das and Madhusudan, 2024).

Despite these limitations, many students surveyed at Warwick use these genAI models to help with their assignments, to produce code or to act as a "study buddy", with most students relying on GPT-3.5 at the time. Some students do critically evaluate the outputs, whilst others do not, with staff reporting an increase in genAI misuse.

3. Performance of AI on university-level work

3.1 Methodology

We collected 122 assignment questions from mathematics and statistics lecturers, who submitted questions from their modules across Years 1 to 4 (FHEQ Levels 4 to 6). The questions were presented to GPT-4o with a *zero-shot* approach, i.e. the AI received no additional guidance or prompting beyond the wording in each question. We classified each question into one of two types:

Proof type. This includes questions that ask for a chain of logical reasoning, often using previous lemmas or theorems. This type of question typically requires little numerical calculations. Examples:

- (Y1) Prove that the composition of two bijective functions is bijective.
- (Y2) Show that the partition function $p(n)$ satisfies a given recursive inequality.
- (Y3/4) Prove that a given Lie algebra is semisimple.

- ***Applied type.*** This includes questions that ask for a concept to be applied to a specific situation, requiring some symbolic manipulation or numerical calculations. The answer is typically a concrete expression, a number, a graph or code.
  Examples:
  - (Y1) Find a particular integral for a given ODE.
  - (Y2) Calculate the first three terms in the asymptotic series of a given integral.
  - (Y3/4) Suggest a proposal density for rejection sampling from a given bivariate distribution. Verify your answer by implementing it in R.

We performed the proof/applied classification to test the hypothesis that genAI is prone to making computational errors in applied-type questions, and less likely to make mistakes in proof-type questions, where the answers are more likely to be in the training data. The split between proof and applied types is shown in Table 1.

Table 1. The distribution of the 122 questions we tested by Year (1, 2, 3/4) and by type (proof, applied).

	Year 1	Year 2	Years 3/4
Proof	27	16	21
Applied	35	7	16
Total	62	23	37

We rated the correctness of GPT-4o's answers on a three-tier (traffic light) scale, where:

- **Green** (70%-100%) indicates a good solution. If produced by a student, it would demonstrate a good understanding of the topic, possibly with a few errors.
- **Yellow** (35%-69%) indicates an adequate or passable solution. If produced by a student, it would show a fair or satisfactory understanding of the topic, with some errors.
- **Red** (0%-34%) signifies a poor solution. If produced by a student, it would indicate a lack of understanding of the topic with fundamental errors.

This scale allows us to quickly analyse questions from a wide range of topics. This system also allows us to obtain an aggregate (expected score) for each year by giving each question the mean score in each category, i.e.

Expected score in each year = $\frac{85 \times N_{green} + 52 \times N_{yellow} + 17 \times N_{red}}{N_{green} + N_{yellow} + N_{red}}$,

where N_i is the number of questions judged to be in category i.

In addition, the lecturers who submitted the questions were asked to re-mark a sample of 35 out of the 122 responses (approximately 30%) and rate them in terms of correctness and in three additional metrics:

- ***Similarity*** to student work (0-100%): A high score means the AI-generated solution closely resembles a typical student submission.
- ***Detectability*** as AI-generated (0-100%): A high score means the solution can be easily identified as AI-generated.
- ***Adaptability*** into student work (0-100%): A high score means the AI-generated output can easily be modified into what appears to be a genuine piece of student work.

The sample was chosen to cover a range of levels and assessment types, and the size was selected so that lecturers were not overburdened with additional work.

3.2 Results

Correctness

The performance of GPT-4o is shown in Figure 1. We see that it performed well on Year 1 assignments, achieving a first-class score. For Years 2 to 4, the performance declined. Lecturers noted that answers to proof questions were often vague, lacked detailed reasoning, or contained significant errors. The AI also struggled with complex multi-step logical arguments. The performance was not uniformly good. Overall, the performance of GPT-4o was comparable to an undergraduate at a mid 2:2 level.

	Year 1	Year 2	Year 3/4
Expected Score	74%	48%	56%

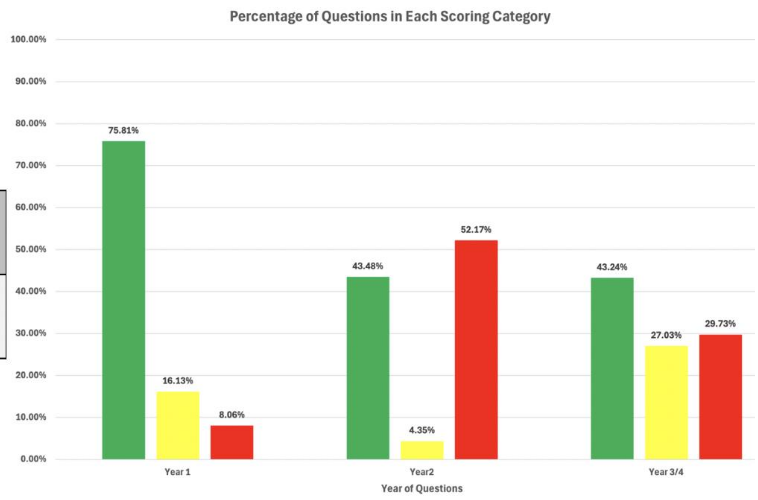

Figure 1. The average score of GPT-4o's answers across various years, and the correctness of the answers evaluated on a traffic-light scale.

Figure 2- shows a performance of proof vs. applied questions. The table shows broadly similarly performance across proof- and applied-type questions. This suggests limited evidence that GPT-4o is better at proof rather than applied questions.

Expected Score	Year 1	Year 2	Year 3/4
Proof	79%	55%	53%
Applied	74%	47%	56%

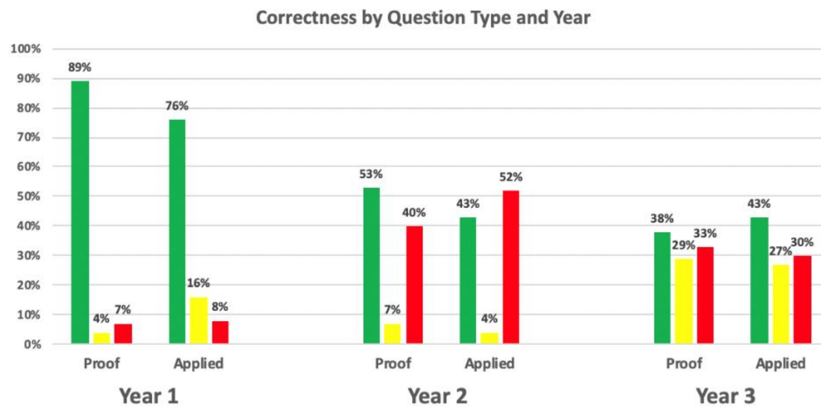

Figure 2. The correctness for proof and applied questions across all years. The scores lend weak support for the hypothesis that GPT-4o is better at proof-type than applied-type questions.

Similarity and Detectability

The similarity score (averaged across all questions in all years) is 62% (see Figure 3), although the distribution is wide. The responses indicated no significant differences between answers to proof-type and applied-type questions.

The detectability score is 53%, signifying some ambiguity in the authorship, again with negligible difference between proof-type and applied-type questions. Lecturers observed that AI-generated responses sometimes included unusual phrasing, excessive verbosity, or atypical grammar — features that could indicate AI authorship.

Adaptability

The adaptability score is 77% (see Figure 3), indicating that answers with AI characteristics could be easily modified by students to resemble their own writing style, e.g. by correcting obvious errors, adjusting the language, and removing AI tell-tale signs.

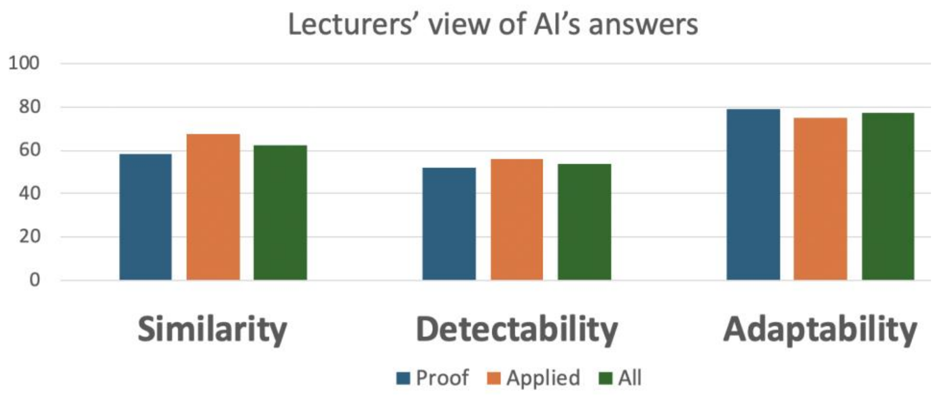

Figure 3. Lecturers' view of GPT-4o's answers, judged in terms of similarity to student work, detectability as AI, and adaptability into student work. The scores are averaged across all years.

These results highlight the nuanced capabilities of GPT-4o. While it demonstrates strong performance on simpler tasks, its limitations in complex reasoning do not necessarily prevent potential misuse if complemented by student critical evaluation of the outputs.

4. How students use genAI.

4.1 Methodology

An online survey of 145 mathematics and statistics students was conducted in June 2024 to assess their use of AI tools, ethical considerations and their attitudes towards AI. This sample represents approximately 7% of the UG and PGT population. Those completing the survey could opt in to a focus group. The respondents consisted of 86 (59%) that declared themselves AI-users (i.e. have used AI tools like ChatGPT for university work) and 59 (41%) non-AI-users.

From those that opted in, a random sample were selected for two focus groups of 6 individuals each. One group comprised AI users and the other non-AI-users.

4.2 Survey outcomes

The survey covered the following areas: ethical considerations, academic integrity, impact on degree value, student attitudes, AI assessment integration, usage patterns, and future concerns.

Three figures on the following pages summarise the survey outcomes.

Figure 4 provides a summary of questions about the students' attitude towards AI, with responses on a Likert scale (strongly disagree to strongly agree), the total number of responses in each category and their respective percentages.

Figure 5 provides a summary of responses regarding frequency of AI use. Figure 6 shows the choice of AI (if any) used by the participants, with ChatGPT being the most popular.

From these results, we made the following general observations (Chongchitnan et al., 2024).

- **Perception of cheating.** Most students regard using AI as cheating, even amongst those who have used AI in assignments.
- **Support for AI-proofing measures.** There is support for proactive measures to mitigate AI misuse, although the effectiveness of such strategies was challenged.
- **Scepticism towards AI accuracy.** Students believe that AI often provides incorrect answers to mathematics and statistics questions.
- **Apprehension about AI's role in future careers.** Students worry that AI might devalue employable skills or make them obsolete.
- **Resistance to shifting assessment methods.** Students are opposed to moving entirely to in-person exams and removing assignments altogether. This suggests a preference for maintaining a mix of assessment methods, highlighting the value students place on assignments as part of their university education.
- **Uncertainty about AI integration.** There was widespread ambivalence about the use of AI in assignments. This uncertainty was shared almost equally between AI users and non-users, suggesting that even those familiar with AI tools remain unsure about the appropriate role of AI in higher education.
- **Ethical concerns.** Some students, particularly non-AI users, refrain from using AI tools due to ethical concerns, such as the fear of cheating or undermining academic integrity. This hesitancy highlights the importance of establishing clear guidelines and educating students on the ethical use of AI in academic settings.
- **Diverse usage patterns among AI users.** While some students use AI tools regularly for assignments, the majority use them sparingly, often for specific tasks like coding assistance

or clarifying concepts. This suggests that AI is being integrated into student work more as a supplementary tool rather than a primary resource.

These findings demonstrate that students, regardless of their personal use of AI, are acutely aware of, and concerned about, the ethical implications of AI in education. It is also interesting to contrast the results in Section 3.2 (GPT-4o's performance) with student perceptions: Whilst GPT-4o can produce accurate and inaccurate responses, only those able to critical evaluate these responses can judge their value and gain educational benefit from genAI.

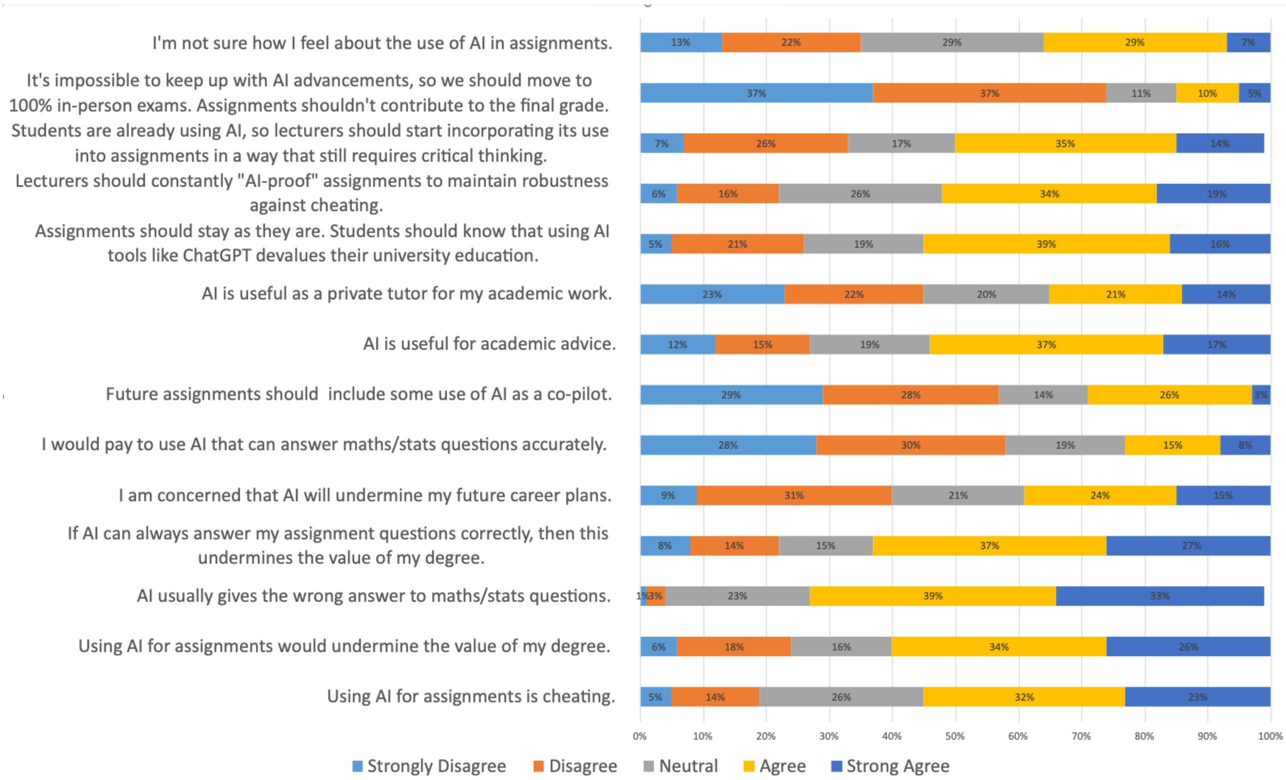

Figure 4. Survey questions with Likert-scale responses. The numbers indicate the percentages in each category.

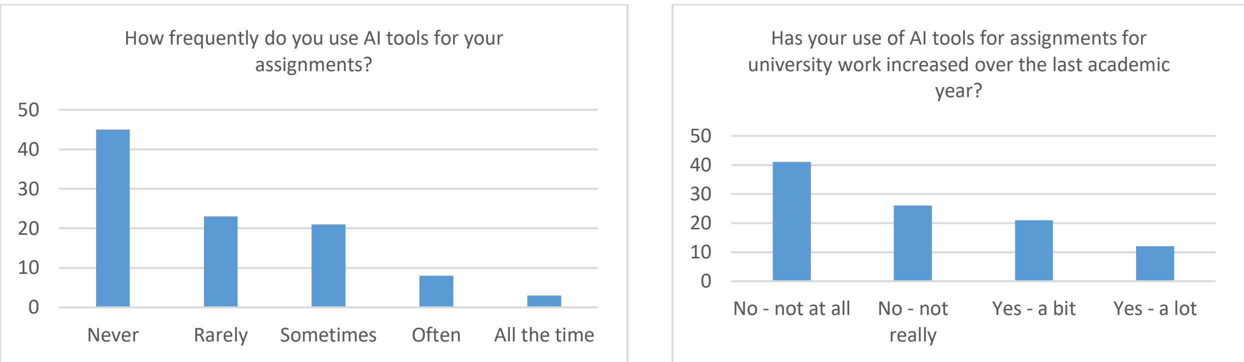

Figure 5. (Left) Percentage responses to the question "*How frequently do you use AI tools for your assignments*?" (Right) Percentage responses to the question "*Has your use of AI tools for assignments for university work increased over the last academic year*?"

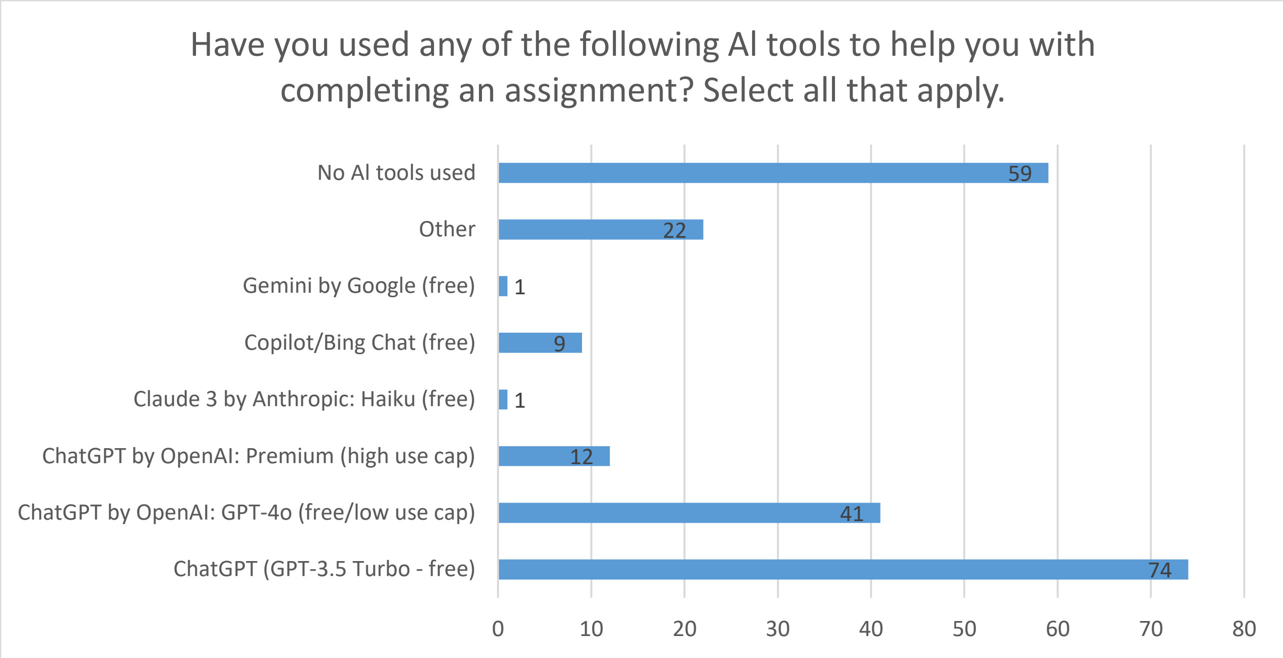

Figure 6. Summary of responses to the question "*Have you used any of the following AI tools to help you with completing an assignment? Select all that apply*."

4.4 Focus-group outcomes

This section presents key insights from focus group discussions conducted separately with AI-users and non-AI-users. The discussions aimed to capture perspectives on how genAI tools like ChatGPT are impacting learning experiences, academic integrity, and future career preparedness.

The discussion is broken down into five thematic areas: Experiences and Attitudes Towards AI, Ethical Considerations and Academic Integrity, Impact on Learning and Skills Development, The Future of AI in Education: Hopes and Fears, and Recommendations for AI Integration. Table 2 provides a summary of the key insights from these discussions. A full analysis is presented in the main report (Chongchitnan et al., 2024).

The focus groups provide individual student thoughts. For example, those who have used AI tools appreciate the support these technologies offer in studying complex concepts.

> *"One thing that* [genAI] *has an edge over asking your professors is the ability to clarify things that you don't really understand in the moment. For example, when I've been reading through my lecture notes and noticed a contradiction, I can scrutinise ChatGPT's answer line by line and ask it again, like why is it doing this?"* — Student D (AI user)

> "*I think it's served as quite a useful tool to replace Googling things... ChatGPT maybe gives you a method that helps you find* [the answers] *a little bit faster... it'll give you that little tip you need in the question to get to the next part.*" — Student C (AI user)

Some students highlighted genAI's limitations in handling advanced problems. Most AI users initially used AI models like GPT-3.5, which has shaped student views. Our work used the more advanced GPT-4o which, although better, showed inconsistent performance.

"*It's hilariously bad at maths. It's very rarely provided anything more useful than just guessing and checking.*" — Student F (AI user)

"*It's hopeless at answering any of my assignment questions.*" — Student A (AI user)

There are interesting contrasts when considering academic integrity. For example,

"*I don't really think I can consider it cheating per se because it just doesn't really give you answers*." — Student C (AI user)

"*People who are going to cheat, they're going to cheat... it's just another tool that's out there.*" — Student 4 (non-AI user)

In terms of learning experience, some provided examples of how they had used AI as a study support tool.

"*It's good for revision plans... it gave me a balance for the 11 exams that I had and helped me prepare for it*." — Student A (AI user)

Others recognised the potential for isolation.

"*I think it could be really detrimental in the fact that it cuts out that communication or that working together aspect of the degree*." — Student 5 (non-AI user)

There were interesting comments regarding potential future usage, both positive and negative.

"*I think AI will become more of a personal assistant/personal tutor that's essentially 24/7 available*." — Student C (AI user)

"*Encouraging people to use AI in their learning promotes bad habits and laziness*." — Student F (AI user)

"*If it reaches the point where it is doing our assignments... then what is the point in our degree at all?*" — Student 4 (non-AI user)

"I think it would create a situation where only students from really high wealth backgrounds are able to access that and then they'd have an extra leg up." — Student 6 (non-AI user)

Table 2. Summary of key insights from the thematic areas identified during focus group discussion.

Area	Discussion area	Key insight
Experiences and attitudes towards AI	Students' initial reactions to AI in academic settings, frequency of AI use, and overall attitudes towards incorporating AI into assignments.	AI users found AI tools helpful for coding and clarifying concepts, while non-users expressed scepticism about AI's reliability and were concerned about its potential to undermine learning.
Ethical considerations and academic integrity	Students' views on the ethical implications of using AI in assignments and whether they perceive AI use as cheating and how this perception differs between AI users and non-users.	Non-AI users largely view AI use in assignments as cheating, expressing concerns about fairness and academic integrity. AI users see it as a tool for assistance rather than a means to cheat.
Impact on learning and skill development	How AI usage affects students' learning processes and skill development, considering the benefits and potential drawbacks of AI in supporting academic growth.	AI users reported that AI helps them understand complex ideas and save time, but they also acknowledge the risk of over-reliance and encountering misinformation.
The Future of AI in Education: Hopes and Fears	Students' perspectives on the future integration of AI in education, including how AI could enhance learning and their fears about potential negative impacts on their degrees and careers.	Students are concerned that AI could devalue degrees and reduce the need for critical thinking, but they also see potential for AI to personalise learning and assist with routine tasks.
Recommendat-ons for AI Integration	Recommendations from students on how AI could be integrated into education including suggestions for guidelines, policy development, and educational practices.	Students would like clear guidelines on AI use, equitable access to AI tools, and assignments that still demand critical thinking and problem-solving skills.

5. Conclusion and discussion

The findings from this study emphasise the need for MSOR educators to develop assessment strategies and policies in response to the rapid development of genAI. We make five recommendations and suggest potential implementations.

1. **Acceptance**. AI is an integral part of the educational landscape, and entirely ‘AI-proofing’ assessments is not feasible. Although some advocate for 100% controlled-conditions assessments, this does not seem feasible. MSOR needs to create AI-ready graduates. Working with AI will involve acknowledging its capabilities and limitations, and integrating it into learning in a manner that enhances education while maintaining academic integrity, for example, when used as a study buddy or for giving additional feedback (Meyer, 2024)

2. **Assessments strategies** should be developed collaboratively with educators and students, fostering innovation and ownership to develop shared ownership of AI potential in MSOR.
3. **Demystifying AI**. Whilst most universities have drawn up generic AI policies, the MSOR discipline has unique characteristics (QAA, 2023). Departments should work with students and staff to clarify the usage policy of AI specifically in MSOR, and educate those who may feel ambivalent about using AI on its benefits and ethical usage
4. **Open dialogue and collaboration.** Encouraging conversations among students, staff and administrators could help address concerns and misconceptions about AI. Co-creation projects and collaborative work could help keep pace with technological advancements, student attitudes and evolving academic practices in the MSOR sector.
5. **Professional development.** The introduction of GPT-o1 which specialises in solving mathematical problems and the anticipated arrival of GPT-5 highlight the need for proactive approaches to maintain the quality and relevance of mathematical assessment in higher education.

GenAI provides new opportunities for innovation and to co-create initiatives where both students and lecturers engage in learning about AI tools together. The keen interest from both staff and students provides a strong opportunity to jointly critically evaluate AI in various ways.

Example 1. Students can learn to verify the accuracy and reliability of genAI in academic work. These initiatives are likely to be more formative than summative and could become part of small-group tutorial work. These sessions should encourage participants to take ownership of their learning by critically assessing AI outputs, understanding the implications of AI-generated content, and discussing the ethical responsibilities associated with AI use.

Example 2. Students create instructions on how to effectively use AI for academic tasks. This may include using AI for summarising notes, finding quotes, creating personalised learning experiences, understanding complex topics and compiling revision schedules.

Example 3. There is the opportunity to examine how assessments can be structured so that AI usage and critical evaluation is encouraged. For example, genAI can be used to generate variations of a proof of a theorem or produce a statistical analysis of a data set. Educators can use these to demonstrate and develop students' ability to critique work. This approach could provide new forms of critique-based assessments.

Example 4. Providing clear examples of acceptable and beneficial AI use. Such examples will need to be tailored to specific modules or learning contexts. For example, it may be appropriate for students to disclose their use of AI in assignments if they rely heavily on AI-generated content or include it directly in their work. In such cases, they should cite the AI tool as they would any other source. Chat logs could form part of an assessment that demonstrates critical engagement with AI. The outcomes of challenges to the outputs can be regarded as evidence of honest and transparent usage of genAI.

In conclusion, we recommend that a proactive and collaborative approach is needed to ensure that educational practices in MSOR subjects evolve in step with the rapid advancement of genAI. While some advocate for a return to fully in-person examinations, we suggest a more balanced approach that leverages the opportunities of this technology to enhance learning and to better prepare students for an AI-enhanced future.

The emergence of genAI in the higher-educational landscape gives rise to new collaboration opportunities between students and staff to engage in a meaningful dialogue. Through this dialogue, sustainable and creative strategies for AI integration might be collaboratively developed. By embracing the potential of genAI, while remaining vigilant about its challenges, we can enhance the MSOR educational experience and prepare students for a future where AI plays a significant role in professional and academic environments.

6. Acknowledgement

This work was supported by the Warwick International Higher Education Academy (WIHEA).

7. Appendix

Example year 1 question and response

We present an example question, GPT-4o input prompt and output.

The following question is a typical Year 1 statistics question asked in a Term 1 probability module.

Question. A random experiment consists of rolling three fair six-sided dice (with face values from the set $\{1,2,3,4,5,6\}$. If two or more dice show the same highest number then the three dice are rolled again. For example, if the numbers on dice are $(5,5,2)$ or $(4,4,4)$ then the three dice are rolled again. However, if the numbers shown are $(5,2,2)$ or $(6,4,4)$ then the dice show a unique highest value and the dice are not rolled again. Each roll of the three dice is called a trial.

- You may assume that all trials are mutually independent.
- You may also assume the independence of the events $\{A_i, i \in N\}$ implies the independence of the events $\{A_i^c, i \in N\}$. (You were asked to show this for three events in the lecture notes.)
- In your answers, you must make clear where you are using independence.

1 Determine the probability that the first trial results in a unique highest number.

2 Let $n \in N$ be fixed. Determine the probability each of the first n trials results in no unique highest number.

3 Let the random variable X denote the number of trials until a unique highest number is obtained. That is, if the unique highest number is obtained on the nth trial then $X = n$. Determine the probability mass function of X.

4 Let $n \in N$. Determine the probability that, if the unique highest number was obtained on the nth trial, then in at least one trial before the nth trial all the dice showed the same value.

Example input prompt. The question was split into separate parts and the question stem repeated. This approach was necessary since on occasion the question context was forgotten between parts. We illustrate the input prompt used for part 4 which is presented verbatim. GPT-4o interprets the LaTeX code correctly.

A random experiment consists of rolling three fair six-sided dice with face values from the set {1, 2, 3, 4, 5, 6}.) If two or more dice show the same highest number then the three dice are rolled again. For example, if the numbers on dice are (5, 5, 2) or (4, 4, 4), then the three dice are rolled again. However, if the numbers shown are (5, 2, 2) or (6, 4, 4), then the dice show a unique highest value and the dice are not rolled again. Each roll of the three dice is called a trial.

You may assume that all trials are mutually independent. You may also assume the independence of the events \{A_i, i \in N\} implies the independence of the events \{A_i^c, i \in N\}. In your answers, you must make clear where you are using independence.

Question: Let n \in N. Determine the probability that, if the unique highest number was obtain on the nth trial, then in at least one trial before the nth trial all the dice showed the same value.

Output. The output is a combination of Markdown and LaTeX. This output has been converted to Word and is presented below.

To determine the probability mass function (PMF) of the random variable X, which denotes the number of trials until a unique highest number is obtained, we need to utilize the fact that each trial is independent and follows a geometric distribution.

First, let's recap the key probabilities calculated earlier:

- The probability of obtaining a unique highest number in a single trial is $p = \frac{55}{72}$.
- The probability of not obtaining a unique highest number in a single trial is $1 - p = \frac{17}{72}$.

1 Geometric Distribution

The random variable X follows a geometric distribution because we are dealing with independent trials and are interested in the number of trials required to get the first success (i.e., obtaining a unique highest number).

The PMF of a geometric random variable X with success probability p is given by:

$$P(X = n) = (1 - p)^{n-1}p$$

where n is the number of trials until the first success.

2 Applying to our Specific Case

Here, the success probability p is $\frac{55}{72}$. Thus, the PMF for X is:

$$P(X = n) = \left(\frac{17}{72}\right)^{n-1}\left(\frac{55}{72}\right)$$

for $n = 1,2,3, \ldots$

3 Final Probability Mass Function

The probability mass function of the random variable X is:

$$P(X = n) = \left(\frac{17}{72}\right)^{n-1}\left(\frac{55}{72}\right), n = 1,2,3, \ldots$$

Thus, the PMF of X can be written as:

$$P(X = n) = \left(\frac{17}{72}\right)^{n-1}\left(\frac{55}{72}\right), n = 1,2,3, \ldots$$

This result uses the fact that each trial is independent, and the geometric distribution models the number of trials until the first success.

8. References

Attewell, S. Student perceptions of generative AI. Jisc report, 2024. Available at: https://edin.ac/3Y8Y2Cd. [Accessed 10 October 2024].

Chongchitnan, S. Parker, M. Mahal, M. and Petrie, S., 2024. Exploring the use of AI in mathematics and statistics assessments, Warwick International Higher Education Academy. Available at: https://warwick.ac.uk/fac/cross_fac/academy/funding/2023-24-int-projects/ai-in-maths/ [Accessed 10 October 2024].

Das, S.R., J.V. Madhusudan, J.V., Perceptions of higher education students towards ChatGPT usage. Int. J. Technol. Educ. 7(1) (2024) 86-106.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., and Liu, T., 2023. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. ArXiv. Available at https://doi.org/10.48550/arXiv.2311.05232

OpenAI., 2022. Introducing ChatGPT. Available at https://openai.com/index/chatgpt/ [Accessed 25 September 2024]

Iannone, P. and Simpson, A., 2022. How we assess mathematics degrees: the summative assessment diet a decade on, *Teaching Mathematics and its Applications: An International Journal of the IMA*, Volume 41, Issue 1, March 2022, Pages 22–31, https://doi.org/10.1093/teamat/hrab007

Iannone, P. and Simpson, A. 2011. The summative assessment diet: how we assess in mathematics degrees. *Teach. Math. Its Appl.*, 30, 186–196.

Iannone, P. and Simpson, A. (eds) 2012. *Mapping University Mathematics Assessment Practices*. Norwich: University of East Anglia.

Meyer, J. et al. 2024. Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions, *Computers and Education: Artificial Intelligence,* Volume 6, (2024), 100199.

QAA, 2023. Subject Benchmark Statement: Mathematics, Statistics and Operational Research. Available at https://www.qaa.ac.uk/docs/qaa/sbs/sbs-mathematics-statistics-and-operational-research-23.pdf?sfvrsn=5c71a881_12 [Accessed October 2024]

CASE STUDY

Student use of large language model artificial intelligence on a history of mathematics module

Isobel Falconer, School of Mathematics & Statistics, University of St Andrews, St Andrews, UK.
Email: ijf3@st-andrews.ac.uk

Abstract

This case study assesses experience in autumn 2023 of permitting the use of Large Language Model Artificial Intelligence (AI) in preparing essays on a module in the history of mathematics. As a check on usage and to ensure academic standards, students were required to complete two paragraphs to accompany their essays explaining their use of AI. These generated qualitative and quantitative data on student familiarity with AI, and ability to use it in a thoughtful and ethical manner, which is reported here. Findings were that over 50% of students rejected AI use, and only 9% used it extensively. There was a weak negative correlation between AI use and essay grade, for which student confidence may have been a confounding factor. The most frequent reasons for rejecting AI were ethical, personal (satisfaction and confidence), and the time needed to correct it.

Keywords: artificial intelligence, generative AI, ChatGPT, student essays, history of mathematics.

1. Introduction

This case study assesses experience in autumn 2023 of permitting the use of LLM AI (Large Language Model Artificial Intelligence, also known as Generative AI) on a module in the history of mathematics. It may help to inform the ‘escalating scholarly interest in AI’s role within educational contexts’ (Bukar et al, 2024). However, reviews of this burgeoning literature note that the majority of studies are theoretical, and that understanding of why students adopt LLM AI, and how they engage with it, is still very limited (Schei et al 2024; Abbas et al 2024).

‘Topics in the History of Mathematics’, is an optional module for mathematics undergraduates in their final or penultimate year at a Scottish university, worth 15 credits (1/8 of their work for the year). Students typically enter the university as high achievers: standard entry grades are Scottish Highers: AAAAB, including A in Mathematics, or GCE A levels: A*A*A, including A* in Mathematics. The student body is roughly split into one third Scottish, one third from the rest of the UK, and one third international students. The two main motivations for taking the module are a desire to broaden their knowledge of mathematics, and a desire to acquire soft and communication skills that are less commonly fostered in core mathematics modules.

The module is assessed through two class tests (totalling 50% of grade), a preliminary essay plan (5% of grade), and an end-of-semester essay on a history of mathematics topic of the student’s choice (45% of grade). This case study covers the essay component.

Our university policy is that use of LLM AI counts as academic misconduct unless a module gives explicit permission for its use. In the 2023-24 module presentation we decided to give such explicit permission for LLM AI use, while taking precautions to ensure that academic standards were maintained. After consulting the Director of Teaching and the University’s LLM AI Guidance, the module team decided to require that essays be accompanied by:

- A paragraph evaluating the ways in which the student had used/not used LLM AI and explaining their decisions. The intention was to be even-handed by asking that all students justify their decisions, whether or not they decided to use LLM AI;
- A paragraph identifying the three most significant sources cited in their essay and what these had contributed to their argument. This paragraph acted as a check that they did, indeed, understand both the argument they had presented, and were familiar with at least some of their sources.

The wording of the assessment rubric relating to LLM AI was iteratively discussed between the module team and the Director of Teaching, and followed closely the University's LLM AI Guidance. The marking criteria were changed from previous presentations to put more weight on quality of argument (which AI is poor at), up to 50% from 40%, and less weight on presentation (which AI is good at), down to 10% from 20%. Students were not directed at any particular LLM AI and were not required to specify what they had used. Appendix A contains the rubric provided to the students in the Module Handbook.

We took this approach for two reasons:

1. Pragmatically, we would be unlikely to detect LLM AI use (Perkins, 2023), so hoped to avoid unknowingly awarding marks for AI-generated content.
2. Pedagogically, use of LLM AI is likely to be part of students' future employment practices so we wished to encourage critical awareness. O'Dea et al (2024, p2) report that 'globally over 43% of employees have used ChatGPT and other large language models, such as Google Gemini and Copilot to help them with their daily work'.

The module team viewed this as a relatively minor change in design of an assessment primarily aimed at evaluating students' history of mathematics skills, rather than as pedagogical research. Hence no ethics clearance was sought; this precludes quoting directly from the wealth of information on student familiarity with LLM AI, and their ability to use it in a thoughtful and ethical manner, that resulted and that is discussed in the remainder of this report.

2. Quantitative data: generation, analysis and results

Sixty-one students submitted essays. Of these, all completed the AI paragraph, and 60 completed the sources paragraph.

Essay marking was split between two markers. One marker assigned grades to the essays before reading the two AI-related paragraphs, and then occasionally adjusted the grade in the light of the AI paragraphs if:

- the paragraphs revealed significant discrepancies between declared AI use and the evidence of the essay, for example if a student identified insignificant, rather than significant, cited sources, or appeared unaware of what the presented argument was, as judged by the second of the required additional paragraphs (none did);
- credit had been given for features of the essay that turned out to be straightforwardly generated by AI as described in the first of the additional paragraphs (two scripts);
- credit had been denied for features assumed to be an over-long quotation but that turned out to be the student's own work with AI in a partner role (see below for partner role, one script).

Students' self-reported use of LLM AI was roughly grouped and labelled into four categories as shown in Table 1. Appendix B gives paraphrases of statements characteristic of each category.

Table 7. Categories of AI usage, showing category label, category name, number and percentage of students who fell within each category

Category label	Category name	No. students in category (N=61)	% students in category
0	No use	33	54
1	Limited use	13	21
2	Moderate use	10	16
3	Extensive use	5	9

Over 50% of students declared that they had not used AI at all.

With this crude categorization, there was a weak negative correlation (-0.37) between use of AI, and overall grade for the essay, i.e. students who made more use of AI tended to get weaker grades.

Grades on the essay were comparable with previous years. However, for the overall module grade a small downward scaling at the bottom end was implemented to bring the distribution in line with previous years.

Although students had not been required to use any specific LLM AI, those that specified their platform all used some form of ChatGPT: 11 specified ChatGPT but did not give the version, two specified ChatGPT-4, one ChatGPT-3.5, and one ScholarAI (a ChatGPT plugin).

3. Qualitative data: generation, analysis and results

Qualitative data came from the submitted AI paragraphs. The initial stages of a grounded theory approach were used to develop themes (Strauss & Corbin 1990), i.e. the paragraphs were all read, and coded with no pre-conceptions of what would emerge, and codes then grouped into higher level themes. Emergent themes were:

- Self confidence;
- Efficiency;
- Self-identity;
- Partner;
- Critical awareness;
- Ethics.

3.1. Self confidence

Student self-confidence appeared most frequently as a factor in students' decisions about LLM AI use and related to two main areas: 1) Language and writing skills, and 2) Existing familiarity with LLM AI. In both areas students ranged from very confident to very lacking in confidence.

Students who expressed little confidence in their *language and writing skills* used LLM AI at many levels, from help with basic vocabulary and grammar, through structuring of paragraphs, to overall structure of the essay and argument (see example comments in §7.4). Many, but far from all, of those seeking help with vocabulary and grammar were students with English as an additional language. But students across the language spectrum used AI to help with structuring at paragraph or overall essay level, pointing out that as mathematics students they had little experience of such tasks.

Conversely, a number of students expressed absolute confidence (sometimes misplaced!) in their ability to write a high-quality essay without AI assistance (see comment in §7.1).

Some students chose not to use AI, as previous *familiarity* led them to believe that it would be of little use in this instance. More frequently, though, students claimed no previous experience and lacked confidence in their ability to instruct it effectively or to evaluate the quality of the result; they chose not to use it for these reasons (see comment in §7.1).

3.2. Efficiency

Students were split on whether using AI would save time and effort and made decisions on this basis. The main time-saving activities mentioned were:

- discovery of sources;
- summarizing sources to build up knowledge and understanding;
- summarizing sources into a literature review.

(see example comments in §7.2, §7.3 and §7.4)

However, such students were outnumbered by those who thought that fact-checking AI-generated research would take more time than it was worth (comments in §7.1). For the majority, this belief was based purely on the number of dire warnings they had read. This was particularly the case for students who had chosen fairly niche topics and were already familiar with the few extant sources; they distrusted what AI might provide if it deemed these not sufficient. Less commonly, anxiety about fact-checking effort came from experience and up-to-date knowledge, such as of a recent rise in LLM hallucinations (a false or fabricated output).

3.3. Self-identity

This theme encompasses factors clustered around students' sense of their own individuality, personal development and satisfaction. Students expressing these views fell almost entirely into the 'no use' or 'limited use' categories (example comments in §7.1).

Many students felt that the arguments they wanted to make were individual to them. They noted that AI-generated writing tended to generalize and to read as generic, whereas the students wanted to make their own precise and detailed arguments, in their own individual style, aimed at a particular audience.

Even where this was not the case, many students felt that doing all the research and writing themselves would improve their own research skills more than using AI would, and that they would derive more personal satisfaction from doing this.

3.4. Partner

Some students used AI in much the way they might use a peer, mentor or supervisor, to bounce ideas off, and check their understanding and interpretation of their sources. Examples included iteratively:

- refining from initial area(s) of interest, or a brainstorm of ideas, into a well-defined essay topic;

- checking and refining the students' translation of material from other languages. This applied not only to students with sources in their own native languages, but also to native English-speakers using sources, for example in French or German. Indeed, there seemed a slightly greater willingness among English speakers to use other-language sources than in previous years (three compared to zero previously);
- helping with understanding and presenting of proofs in unfamiliar areas of mathematics, for example, by describing the proof in simpler/more modern English, by explaining the reasoning behind the steps, or by assessing the accuracy of a student's account of the proof against the original proof.

(see example comments in §7.4)

Note that the effectiveness of these uses, judged by quality of outcome, has not been evaluated.

3.5. Critical awareness

Students developed their own critical awareness of LLM AI in two ways: through external reading, and through trial. External reading (sometimes cited) was often used to justify claims that:

- AI would not handle well topics that were very specific with few sources, rendering it more prone to hallucinating;
- AI use is unethical in a variety of ways.

Some students took the opportunity offered by the essay's rubric to trial and experiment with AI, especially if they had little or no previous experience (comment in §7.2). The majority of trials compared the AI output with something they had written independently themselves; they generally reported that AI had missed or distorted the point of their argument, and required so much correction that it was easier just to write their own text. One or two students trialed other aspects of AI, such as comparing its search effectiveness with that of Google Scholar.

3.6. Ethics

Through their reading, many students became aware of ethical issues around using AI. That most frequently raised was around the originality and authenticity of AI-generated work, as it is based on vast quantities of untraced and unacknowledged data (e.g. Chesterman 2024). The associated danger of spreading misinformation was strongly raised by some students (e.g. Xu et al 2023) (comments in §7.1).

Other ethical issues raised (by one student each) were:

- Sustainability – the environmental impact of data centres and of mining/disposing of rare earths for components (see, e.g. Henderson et al 2020);
- Racism (and many other 'isms') as LLM AI is based on historical data and hence traditional stereotypes and patterns of expression (see e.g. Bender et al 2021).

4. Discussion & Conclusions

The module's AI rubric was fairly successful in prompting students to inform themselves about LLM AI and to think critically about its use. Indeed, one or two students interpreted the rubric as meaning that they had to use AI to at least some extent, and they trialed it accordingly.

Having informed themselves, a surprisingly high number rejected its use completely, so it is not clear that any aim of enhancing skills in effective use of AI were realized. However, any such aim was secondary to the main purpose of the assessment to develop history of mathematics skills. The most frequent reasons for rejecting LLM AI were ethical (25%), personal (satisfaction and confidence) (40%), and the time needed to correct it (40%).

A minor positive development was the increased willingness observed among English-speaking students to tackle sources in other languages. The effect was small (three students) but noticeable compared with the complete lack of such students in previous cohorts.

It seems likely that a confounding factor underlying the weak negative correlation between AI usage and grade, was student confidence and ability in written English; students whose written English was weak, as judged by the reasons they gave for using AI and their performance in class tests, were much more likely to be moderate or extensive users of AI.

It is possible that, overall, the students did better than previous cohorts who did not have access to LLM AI. The standard of the best essays seemed very comparable to the best essays of previous cohorts, but these were written by students who had rejected AI. At the lower end of the scale, the need for downward scaling of the overall module grade in order to bring grades into line with those of previous cohorts might indicate that weaker students had benefitted from AI use. However, since scaling analysis is done at module level rather than that of separate assessment components, further analysis would be required to disentangle the essay from the class test results of this and previous cohorts, before comparisons could be made. A qualitative comparison of the corpus of essays from this cohort with those of previous cohorts, could provide insights into how AI affects student writing quality and originality, but would be difficult to report on robustly given the lack of ethics clearance for any of these assessments.

Overall, this intervention proved easy to implement, taking little additional resource in class time or marking; the major time taken was in module team discussions beforehand when developing the rubric. Judging by the good correspondence between the AI statements, the sources paragraphs, and the essays themselves, it appeared that the students were honest in reporting their AI use, suggesting that the intervention was effective in its major aim of making any LLM AI use transparent to the markers; whether its success could be repeated with students who were generally less engaged and motivated, is less clear.

More forethought for the possible value of the additional paragraphs beyond the primary assessment task, might have prompted an application for ethics clearance and enabled more robust reporting of the outcomes to the wider HE mathematics community; seeking ethics clearance would seem advisable if there is even a remote likelihood that outcomes may form the basis for research, however the impact on what they write of asking students for the necessary informed consent has also to be considered.

Although the approach taken was, on balance, a success, we cannot assume that it can be repeated on the next presentation in 2025-26, as AI and student skills with it, will have moved on considerably by then.

5. Acknowledgements

I would like to thank Dr Deborah Kent, who co-lectured this module, and the students on MT4501 who were a joy to work with.

6. Appendix A: Project requirements and marking criteria as stated in the module handbook

Project requirements:

- Free choice of topic – provided it is about history of mathematics;
- An essay, normally of 2500-3000 words (depending a bit on how many equations or tables you use), containing:
    - First page with Title, your student ID, an abstract of 3-5 sentences describing the content of the essay;
    - Introduction, including your research questions and thesis statement;
    - Body of the Essay (may be divided into sections but does not have to be);
    - Conclusion;
    - Citations and references in a standard format (see below);
- PLUS;
    - a compulsory paragraph of up to 200 words evaluating the ways in which you have used/not used AI. If you have used AI say how, and what it contributed to your essay; if you have not, explain your decision not to;
    - a compulsory paragraph of around 200 words that identifies the three most important sources you have used and analyses the ways in which those were important to your argument.

Use of LLM/AI (e.g. ChatGPT)

On MT4501 we recognize the benefits of learning to use LLM/AI effectively and intelligently.

You may use LLM/AI for your project, but we want to know how and why you have used it. If you have not used it, tell us why you decided not to. Either answer is equally acceptable. You **must** submit a paragraph accounting for this along with your project essay. At a minimum, we expect you to have verified all the references and "facts" contained in your essay, and to have chosen an essay structure that provides the most effective support for your argument.

LLMs may be useful for:

- writing your essay for you (!);
- revising your drafts to improve the quality of your English, especially if you are a non-native English speaker;
- structuring your essay (in a common way).

But we expect you to demonstrate awareness of limitations of LLMs such as:

- LLMs may generate misinformation, as they prioritize coherence and plausibility over factual accuracy;
- LLMs generate text that is coherent, contextually relevant, and plausible, but they do not "think" and cannot generate an argument;
- LLMs may generate an essay structure that is common and presents the content in a coherent manner, but this may not be the best structure to support your argument.

Note also the University's guidance and policy on Good Academic Practice has sections on unauthorised use of AI and how to avoid it.

We will discuss use of LLM/AI in a tutorial.

Project marking criteria

Essays will be marked according to the following criteria:

- Quality of argument/analysis (weighted approx. 50%), including:
  - Originality/independence of approach;
  - Difficulty/ambition of project;
  - Critical writing, analysis and interpretation;
  - Understanding of concepts;
  - Were appropriate assumptions made & appropriate conclusions/inferences drawn?;
  - Were appropriate tools/methods used?;
  - Was the argument well-supported by the evidence?;
- Quality of content (weighted approx. 40%), including:
  - Choice of appropriate sources and examples;
  - Amount of work undertaken;
  - Appropriate use of diagrams, tables, images;
  - Factual accuracy;
  - Understanding of detail;
- Presentation and exposition (weighted approx. 10%), including:
  - Statement of aims and objectives;
  - Structure and organization of material;
  - Clarity and readability;
  - Literacy and grammar;
  - Citation and referencing.

The two compulsory additional paragraphs will be used to assess your essays against the standard School grade descriptors.

The paragraph on your use/non-use of LLM/AI will be assessed according to:

- Depth of reflection on effective ways to use LLMs, and their limitations;
- Correspondence between your use/non-use of LLMs and the evidence of your essay.

The paragraph on your most important three sources will be assessed according to:

- Appropriateness of your selection of sources to discuss;
- Quality of your argument about what they have contributed to your essay.

7. Appendix B: Characteristics of AI use categories

Paraphrases (not direct quotations) of AI statements characteristic of each category.

7.1. No use

'I chose not to make use of AI.' Such statements were often followed by reasons such as:

> 'I have not used AI before and did not want to spend time learning to use it effectively, rather than researching for my essay.'
>
> 'There are very few sources on my specific topic and I was worried that AI would generate false sources to fill the gaps.'
>
> 'Fact-checking every AI output would take more time than writing the essay myself.'
>
> 'I felt confident in my ability to write clearly and concisely, and doing so would give me more personal satisfaction.'
>
> 'I could not justify using AI for an academic essay, due to its environmental costs.'
>
> 'AI relies on the creative work of individuals who are not acknowledged or paid; I could not use it in good conscience.'

7.2. Limited use

'I did not use LLM/AI for very much.' Such statements were usually followed by an account of trials they had performed with AI on their own initiative to assess its usefulness, finding it not useful, for example:

> 'I tested ChatGPT by asking it to summarise this source, but the summary was over-simplified and omitted key points, and the writing felt impersonal.'

'I used AI to provide an initial structure to help me get started, but abandoned the structure as my research progressed.'

'I used AI to suggest potential avenues of research, but then checked them out and decided whether to pursue them, researching them on my own. I did not use AI for any of the writing.'

'Once I had written my essay, I used AI minimally to suggest improvements to spelling, grammar and clarity.'

7.3. Moderate use

'I have used AI to help structure my essay, and to improve the quality of my English by fixing the grammar and suggesting more varied word choices.'

'I chose to use AI to help express my points more concisely and reduce my word count. I also used it to help generate a title and abstract.

7.4 . Extensive Use

'I used AI to organise my thoughts and refine my essay plan, and then to help break the plan down to actionable subsections. I used it extensively to check my writing style. Finally, I used it to find extra sources from the internet.'

‘I have used AI throughout my essay. During the research I would ask it questions, and its responses would provide me with the key features of a subject – which I could then check whether I wanted to include. I did not use it for help with writing text, although I did use it to highlight grammatical errors. I also used it to search out answers to technical questions about use of LaTeX.’

‘I used AI extensively, especially to help me understand the unfamiliar style of proofs. I asked ChatGPT to explain the proof, then wrote the proof in my own words and asked ChatGPT to check my proof against the original to make sure I had not omitted key steps or information.’

‘Using AI tools improves quality and efficiency in the essay writing process. It enhances the precision of language, and streamlines discovery and review of the literature.’

8. References

Abbas, M. (2024). Is It Harmful or Helpful? Examining the Causes and Consequences of Generative AI Usage among University Students. *International Journal of Educational Technology in Higher Education* 21(1). https://doi.org/10.1186/s41239-024-00444-7.

Bender, E.M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. New York: Association for Computing Machinery. pp.610-23. https://doi.org/10.1145/3442188.3445922.

Bukar, U.A., Sayeed, M.S., Razak, S.F.A., Yogarayan, S., and Sneesl, R. (2024). Decision-Making Framework for the Utilization of Generative Artificial Intelligence in Education: A Case Study of ChatGPT.” *IEEE Access* 12 pp.95368–89. https://doi.org/10.1109/ACCESS.2024.3425172.

Chesterman, S. (2024). Good Models Borrow, Great Models Steal: Intellectual Property Rights and Generative AI. *Policy and Society*, 00(00), pp.1–15. https://doi.org/10.1093/polsoc/puae006.

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D. and Pineau, J. (2020). Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. *Journal of Machine Learning Research* 21(248) pp.1–43. http://jmlr.org/papers/v21/20-312.html [accessed 23 December 2024].

O’Dea, X., Tsz Kit Ng, D., O’Dea, M., & Shkuratskyy, V. (2024). Factors affecting university students’ generative AI literacy: Evidence and evaluation in the UK and Hong Kong contexts. *Policy Futures in Education*, *0*(0). https://doi.org/10.1177/14782103241287401.

Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice*, 20(2). https://doi.org/10.53761/1.20.02.07.

Schei, O.M., Møgelvang, A., and Ludvigsen, K. (2024). Perceptions and Use of AI Chatbots among Students in Higher Education: A Scoping Review of Empirical Studies. *Education Sciences* 14(8): 922. https://doi.org/10.3390/educsci14080922.

Strauss, A. and Corbin, J.M. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Thousand Oaks, CA, US: Sage.

Xu, D., Fan, S. and Kankanhalli, M. (2023). Combating Misinformation in the Era of Generative AI Models. *Proceedings of the 31st ACM International Conference on Multimedia*. New York: Association for Computing Machinery. pp.9291–98. https://doi.org/10.1145/3581783.3612704.

WORKSHOP REPORT

Designing Assessment to Promote Students' Wellbeing

Noel-Ann Bradshaw, Faculty of Engineering and Science, University of Greenwich, London and Kent, UK. Email: N.Bradshaw@gre.ac.uk
Tony Mann, School of Computing and Mathematical Sciences, University of Greenwich, London and Kent, UK. Email: A.Mann@gre.ac.uk

1. Workshop Report

An in-person workshop on "Designing Assessment to Promote Students' Wellbeing", organised by Noel-Ann Bradshaw and Tony Mann of the University of Greenwich, was held on 10 July 2024 at the University of Greenwich in London, as part of the Higher Education Teaching and Learning Workshop Series 2023/24 jointly offered by the Institute of Mathematics and its Applications (IMA), Royal Statistical Society (RSS) and London Mathematical Society (LMS) (IMA, 2024). The workshop was attended by 39 participants from 16 different universities in the UK and Ireland. This was a follow-up to a workshop held in July 2023 which had provoked discussion about how assessment can affect student's wellbeing and mental health (IMA, 2023): this new workshop set out to explore ways in which assessment design can help mitigate any damaging impact higher education assessment in mathematics can have on students, and to encourage debate around this important topic.

It began with a talk by Gwen Thomas (University of Greenwich) on *Supporting Neurodivergent Students for more effective learning and assessment*. The presentation included a number of activities which vividly illustrated to the audience how different people respond to study situations in different ways. This was followed by a presentation by Noel-Ann Bradshaw (University of Greenwich) on *Perspectives on mathematical assessment from a mature student & Senior Manager*, in which the speaker discussed her own experience as a student. Then Robyn Goldsmith (Lancaster University) spoke on *Building a Student-led Mental Health Community*, covering her experience with a student society she founded to support students' mental health. The morning presentations provided the context and set the tone for the afternoon sessions in which several academics presented examples of their practice.

After lunch, Sue Pawley (Open University), in a talk entitled *Rapid query responses, online mocks & other ways to reduce assessment anxiety*, presented work she had done with Cath Brown on helping students prepare for assessments, and Brendan Masterson, Alison Megeney, and Nick Sharples (Middlesex University), talking about *Authentic, no-exam assessment for student wellbeing*, told us about their innovative approach to assessment. Wodu Majin (University of Sheffield) in a presentation *Easing the burden on memory: Mind map assessments in mathematics* described an unusual assessment she had used to help students structure their understanding of a branch of mathematics, and Tony Mann spoke on *Managing group assessments to minimise the impact on students' wellbeing*, discussing aspects of his practice intended to make groupwork less stressful for students.

The final session of the day consisted of small group discussions, allowing participants to share their experience. The conversations showed how passionately the participants care about their students' wellbeing and how the topic of the workshop resonated with many in the UK maths higher education community. In particular the discussion and feedback from participants indicated a strong feeling that there is a need to reduce the use of time-constrained exams and for degree programmes to offer more opportunities for reflective writing.

The organisers hope that the workshop will help inspire the development of new approaches to assessment in mathematics that will benefit future students. It was especially gratifying for the organisers that the post-workshop survey of participants showed that many of those present indicated that they will be considering adjusting their practice to address some of the issues discussed.

2. Speakers and Abstracts

Gwen Thomas (University of Greenwich): *Supporting Neurodivergent Students for more effective learning and assessment*.

Abstract: A brief look at (and experience of) the challenges that some students face in the learning and assessment environment, and some of the reasonable adjustments that can make a difference.

Noel-Ann Bradshaw (University of Greenwich): *Perspectives on mathematical assessment from a mature student & Senior Manager*.

Abstract: This talk will share a unique perspective as someone who struggled with their mental health during their UG mathematics degree as a mature student and now, 20 years later, has shared responsibility for the outworking of the University assessment policy for STEM subjects at Faculty level.

Robyn Goldsmith (Lancaster University): *Building a Student-led Mental Health Community*.

Abstract: The student-led Mind Society at the University of Greenwich was dedicated to creating an open and safe space for students to talk about mental health. As the founder of the Mind Society when I was an undergraduate studying mathematics, I will bring the student perspective of assessment and wellbeing, sharing experiences of three years of building community, raising awareness of student mental health and abolishing stigma.

Sue Pawley (Open): *Rapid query responses, online mocks & other ways to reduce assessment anxiety*.

Abstract: Students often find completing assessments very stressful but are reticent to seek help and advice. In this presentation I will talk about several support initiatives at The Open University that aim to help reduce student assessment anxiety.

Brendan Masterson, Alison Megeney, Nick Sharples (Middlesex University): *Authentic, no-exam assessment for student wellbeing*.

Abstract: There is compelling evidence that high-stakes exams are detrimental to student wellbeing and further that these effects are not uniform across demographics. The Middlesex maths team will share their experiences of replacing all exams on specialist maths modules with authentic coursework assessments for a better and fairer student experience.

Wodu Majin (University of Sheffield): *Easing the burden on memory: Mind map assessments in mathematics*.

Abstract: In this presentation, I will describe an assignment in which students produced mind maps in a module that heavily featured numerical methods. This assignment emphasised aspects of mathematics that traditional assessments might not directly address. I will reflect on the implementation of the assignment, student engagement with it, and possible psychological benefits of this type of assessment.

Tony Mann (University of Greenwich): *Managing group assessments to minimise the impact on students' wellbeing*.

Abstract: Graduate employers want university mathematics degrees to develop skills in working with people, but groupwork can be very stressful for students. I will present examples from my experience and discuss how I have adapted my practice, including ideas from many colleagues, to try to make student groupwork as valuable as possible while seeking to reduce any pressure it places on students' mental or physical health.

3. Acknowledgements

We are grateful to the IMA, LMS, and RSS for their support of this workshop, to the University of Greenwich for hosting, to the workshop speakers for their presentations and for granting permission to publish their abstracts here, and to all the participants whose contributions to the workshop discussions made this such a productive event.

4. References

IMA, 2023. Assessment in Mathematics and its Effect on Student Wellbeing. Available at: https://ima.org.uk/22451/assessment-in-mathematics-and-its-effect-on-student-wellbeing/ [Accessed 25 July 2024].

IMA, 2024. Designing Assessment to Promote Students Wellbeing. Available at: https://ima.org.uk/24833/designing-assessment-to-promote-students-wellbeing/ [Accessed 23 February 2025].