CLASSROOM NOTE

Beyond traditional grading: introducing specifications grading to a final-year course in mathematical biology

P. Iannone[a], S. O'Hagan[a], and N. Popović[a]

[a] School of Mathematics, The University of Edinburgh, James Clerk Maxwell Building, The King's Buildings, Peter Guthrie Tait Road, Edinburgh EH9 3FD, Scotland, United Kingdom

ARTICLE HISTORY
Compiled August 13, 2025

ABSTRACT
This classroom note describes the redesign of a final-year undergraduate course in mathematical biology at the University of Edinburgh to incorporate specifications grading. Traditional grading approaches in undergraduate mathematics often do not accurately reflect a student's grasp of the material, typically rewarding partial understanding without encouraging holistic comprehension or iterative improvement. To address these shortcomings, we introduced an assignment regime, structured around the EMRN Rubric, which allowed students multiple attempts at written homework, incentivising them to meet high standards and to act directly upon feedback. Our findings suggest that specifications grading led to increased student engagement and collaboration, but also that students found the regime highly unusual compared with their experiences of traditional grading.

KEYWORDS
Assessment; feedback; specifications grading; EMRN Rubric; undergraduate mathematics education

1. Introduction

By way of motivation, we begin by describing what we perceive to be some of the shortcomings of traditional approaches to the grading of continuously assessed student work (coursework) in undergraduate mathematics in the United Kingdom. We then briefly describe the course in question, and its context, before specifying in detail the changes made for the academic year 2023–24 and discussing their implementation and impact.

We are aware that there is no single best approach to teaching and assessment. Hence, in writing this note, we do not claim that our approach is optimal, nor that the changes made in our context will be effective in others. Rather, we wish to share our experiences in the hope that others may find them thought-provoking and possibly even encouraging. We are also keen to learn from practice in other institutions; with that aim, we are hopeful that this note may initiate discussion and collaboration.

CONTACT S. O'Hagan. Email: s.ohagan@ed.ac.uk

1.1. Traditional grading, its shortcomings, and alternatives

In their book ‘Grading for Growth’ (Clark & Talbert, 2023), the authors argue that traditional grading has several shortcomings. They point out that, for example, a grade of 60% often conveys little information about a student’s demonstrated achievements in a course. By ‘traditional grading’ they—and we in this note—mean a practice where students complete various pieces of work, each in a single attempt, that are then awarded a numerical mark by a marker. While feedback may be given, there is no opportunity for a student to act directly upon that feedback and have their improvement recognised or rewarded. With traditional grading, once a student has ‘lost’ marks for a piece of work, those marks are gone forever. It is also typical of traditional grading for partial credit to be given for specific lines of working or particular logical deductions, regardless of the overall quality of a solution.

Students who have come through the Scottish school system—or any of the other education systems in the United Kingdom—have ample experience of traditional grading, as it is pervasive in public mathematics examinations. We believe that this pervasiveness establishes an academic culture where students expect partial credit for partially correct work, regardless of the holistic quality and logical coherence of an attempted solution to a problem. That culture continues into higher education, too, where it is often possible for a student to attain a good passing grade in a course without ever having solved a mathematical problem correctly, in any meaningful sense. Provided a sufficient number of the individual parts or steps in a submission are correct, a student can accumulate sufficient partial credit to pass, or even to score quite highly.

Another shortcoming of traditional grading is that it is somewhat artificial. First, it does not reflect the real-life practice of doing mathematics. Mathematicians will rarely produce a print-ready piece of research without receiving feedback, either informally or through the peer review process. Ideally, we would like to replicate that practice in higher education by offering students experience of mathematics as a collaborative and iterative endeavour. Second, traditional grading bears little resemblance to how effort and attainment are appraised in the workplace.

None of these observations should be interpreted as criticism of students. In the game of assessment and certification, they are not usually the ones setting the rules. In our experience, students are keen to learn and ambitious to achieve. As teachers working in higher education, we have substantial control over the expectations we set and the behaviours we choose to value and reward.

There are, of course, many alternatives to traditional grading in mathematics, with assessment practice varying widely by educational context. Within higher education in the United Kingdom, closed-book examinations remain, by far, the most common method of assessment (Iannone & Simpson, 2022). While a literature search revealed numerous publications on specifications grading in mathematics, we found that most were concerned with school-level or early undergraduate mathematics, and that little has been written about specifications grading in advanced undergraduate mathematics courses.

1.2. Course description and design

Mathematical Biology (MATH10013) is a one-semester (11-week) course offered by the School of Mathematics at the University of Edinburgh in Scotland. It is typically taken by undergraduate students in their fourth (and final) year of study.[1]

In 2023–24, the contact teaching activities in the course comprised two 50-minute lectures each week and a 50-minute workshop every second week. Lectures were 'flipped' (King, 1993), in the sense that students were assigned weekly reading from a collection of undergraduate textbooks; in class, the lecturer expanded on the assigned reading by presenting examples of mathematical models that were either drawn from the reading, or based on it. Only the most fundamental concepts were discussed explicitly by the lecturer in class. The course was assessed through a combination of a written final examination (60%), written homework assignments (25%), and automated online quizzes (15%). The written homework assignments consisted of two problems each, due every two weeks, that were drawn verbatim from problem sheets. While students were encouraged to discuss the problem sheets in workshops, the two assessed problems were not announced until after the workshop preceding the submission deadline, to encourage engagement with all problems on that sheet. The course had an enrolment of 75 students; 71 of those completed the final examination. The teaching team consisted of the lecturer, Nikola Popović, and two graduate teaching assistants (TAs) who assisted with workshops and the marking of assignments. Over the summer of 2023, preceding the start of teaching, the lecturer consulted closely with a learning designer, Steven O'Hagan, to plan and implement a modified assessment structure as described in the following section.

2. Implementation of specifications grading

With the aims as described above, the homework assignment regime was modified for the academic year 2023–24. Submissions were assessed using a form of specifications grading (Nilson, 2014), where high standards are set and students have multiple opportunities to demonstrate that they have met them. In particular, we applied the so-called EMRN Rubric, as written about by Talbert (Talbert, n.d.), which is based on the work of Stutzman and Race (Stutzman & Race, 2004). The rubric is summarised in Figure 1; we chose it because we felt that it would address our perceived shortcomings of traditional grading: students would be incentivised to produce high-quality work and to engage with their markers' feedback. By reducing the process of marking to two binary judgements, we aimed to minimise the time spent by markers on assigning a grade, allowing them to focus on providing constructive feedback instead.

Crucial to our implementation of the EMRN Rubric was reaching a shared understanding among the teaching team of what the lecturer meant for a piece of work to 'meet expectations'. After the first submission deadline, the lecturer, learning designer, and TAs met to discuss the assignment regime and agree standards: they marked several submissions collaboratively and considered edge cases in order to promote consistency and calibrate expectations. Occasionally, throughout the semester, TAs approached the lecturer for clarification and advice when they were unsure of what grade to award particular submissions. However, those instances were rare: consensus was swiftly established on the teaching team that any fundamental mistake in reasoning or logic of argument would be marked as 'Revision Needed' (R), whereas a submission that was fundamentally correct would typically be rated 'Meets Expec-

[1] According to the Scottish Credit and Qualifications Framework (SCQF), the course is worth 10 credit points at level 10. A course description for students, information about prerequisite courses, and summary of learning outcomes for the academic year 2023–24 can be found in the University of Edinburgh's Degree Regulations and Programmes of Study, under http://www.drps.ed.ac.uk/23-24/dpt/cxmath10013.htm.

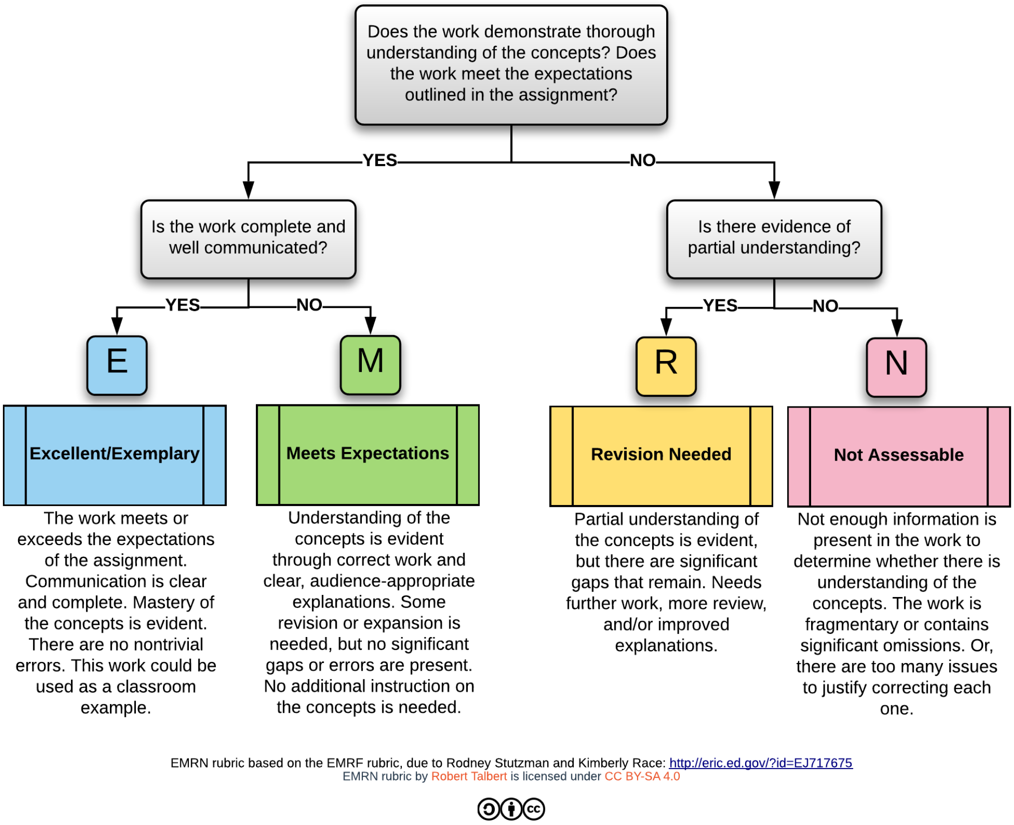

Figure 1. A flow diagram describing the EMRN Rubric.

Table 1. The mapping of letter grades to numerical marks.

Grade	E	M	R	N
Mark	2.5	1.75	0.5	0

tations' (M) even if it contained multiple minor algebraic mistakes. The distinction between 'E' and 'M' was more subtle, and generated the vast majority of queries by TAs to the lecturer, as that is where the inherent subjective differences between individual markers were most evident. Finally, a grade of 'N' was reserved for submissions that were too fragmented to gauge a student's understanding.

The grade descriptors of the EMRN Rubric were intended to provide some clarity to students about expected standards. However, these do contain subjective terms—such as 'correct', 'clear', or 'significant'—that depend on the context of the course and may initially be unclear to students. Given the cohort of mostly final-year mathematics students, we hoped that the class would have an intuitive understanding of what was expected given the descriptors. Model solutions to all unassigned problems from problem sheets were made available to students following the biweekly workshops to exemplify expectations; moreover, particularly high-quality exemplars of student work were shared with the class following each assignment to help students hone their understanding of their markers' expectations throughout the semester. A brief guide to mathematical writing was also made available to the class at the beginning of the semester.

A key feature of the assignment regime was the revision and resubmission of work. Students had the opportunity to act on feedback and resubmit their revised work concurrently with the next homework assignment. Students were encouraged to attend office hours to discuss their feedback with the course lecturer before resubmission. Where a student chose to resubmit, the final mark on that assignment was taken to be the higher of the original mark and the one attained by the resubmission.

Institutional systems required a numerical mark to be returned to students for each component of assessment, in alignment with the University of Edinburgh Common Marking Scheme (CMS). Thus, we had to decide on how to translate grades in the EMRN Rubric to numerical marks. We settled on awarding a mark out of 2.5 for each assigned problem, as shown in Table 1, which implied that the 10 written homework assignments together were worth 25 marks, with one mark being equivalent to one percentage point of course credit. Under that mapping, a student meeting expectations (M) on every assigned problem would score $17.5/25 = 70\%$, which corresponds to an 'excellent' A grade in the CMS. A student not meeting the expected standard on any assignment could score a maximum of $5/25 = 20\%$, which is described as a 'clear fail'. A spreadsheet was shared with students at the beginning of the semester so that they could explore various combinations of grades and the equivalent numerical totals in the CMS.

We have presented the details in the previous paragraph for three reasons: first, to note that we were somewhat constrained by institutional systems, but that we were able to adapt to them; second, to underscore the point that we can set high expectations of students if we choose to; third, because the requirement to report numerical marks did have some negative implications, which we discuss below.

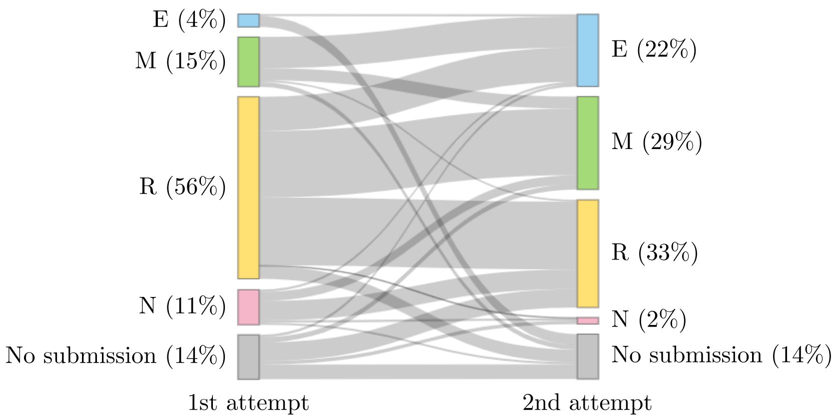

Figure 2. The percentages of grades attained by all 750 first and second attempts at assignment problems.

3. Discussion and conclusions

3.1. Student attainment

Figure 2 shows the grades awarded across all 10 assignment problems in the course.[2] The flow from left to right illustrates how the grade attained by a submission changed between students' first and second attempts. Each column represents 750 submissions—10 problems for each of the 75 students enrolled on the course. Around two thirds (67%) of first attempts were assessed as needing revision (R) or being not assessable (N), with only 19% meeting (M) or exceeding (E) expectations.

There was a high level of engagement with the resubmission process; correspondingly, most of the work that was revised and resubmitted following feedback then met expected standards. Specifically, 93% of first attempts graded as needing revision (R) or being not assessable (N) were revised and resubmitted following feedback, indicating that the class did engage with the process. Of these resubmissions, the majority (52%) then met expectations (M) or exceeded them (E). These figures support the view of the teaching team, discussed below, that students were acting constructively on feedback during the revision stage, with the aim of improving their work.

A significant number (157 of 421; 37%) of first submissions graded as needing revision (R) remained at that grade after resubmission. We observed two main underlying causes: either, the student addressed the given feedback, but introduced different errors at the revision stage; or, there was mismatch between the intended feedback and how it was interpreted by the student.

Figure 3 shows an example of student work, illustrating typical feedback given on a first attempt and how it was acted upon by the student during the revision stage. It is worth noting that the feedback did not have to be extensive in order to help the student improve, provided it made clear where they should focus their attention, thus enabling them to take responsibility for their own learning. Indeed, we share the view expressed by Wiliam (2011) that feedback should create 'more work for the recipient than the donor'.

[2]Ethics approval was obtained through the School of Mathematics at the University of Edinburgh.

Question: We have the Malthusian model where $p(t)$ grows according to

$$\frac{dp}{dt} = rp, \quad \text{with } p(0) = p_0,$$

where r is a birthrate constant. At time t_0 we harvest the population reducing it from $p(t_0)$ to $ap(t_0)$, with $0 < a < 1$. How long does it take for the population to recover to what it was just before harvesting.

Answer: We can solve the differential equation to find that $p(t) = p_0 e^{rt}$. Thus, the population at t_0 is $p_0 e^{rt_0}$ and the population after harvesting is $ap_0 e^{rt_0}$. The time t it takes for the population to recover can thus be found by solving the equation $ap_0 e^{rt_0} e^{rt} = p_0 e^{rt_0}$, which gives $t = -\frac{1}{r} \log a$.

Give more detail.

Explanation missing -- how is that condition obtained?

(a) A student's first attempt at an assignment problem, which was graded as 'Revision Needed' (R).

Answer: We can solve the differential equation using separation of variables:

$$\frac{dp}{p} = r dt \iff \int \frac{dp}{p} = \int r dt \iff \log(p) = rt + C \iff p(t) = e^{rt+C} = e^C e^{rt}.$$

Setting $t = 0$ shows us that $e^C = p_0$. Thus, the population at t_0 is $p_0 e^{rt_0}$ and the population after harvesting is $ap_0 e^{rt_0}$. If we imagine the model restarting after harvesting has happened, the growth will remain the same (i.e. exponential), however our initial condition will now be $p(0) = ap_0 e^{rt_0}$. Thus the population at time t, where t is now the time since harvesting occurred, will be

$$p(t) = ap_0 e^{rt_0} e^{rt}.$$

If we let τ represent the time it takes after harvesting for the population to recover, it follows that

$$p_0 e^{rt_0} = p(\tau) = ap_0 e^{rt_0} e^{r\tau}.$$

We finish by solving this equation for τ explicitly:

$$p_0 e^{rt_0} = ap_0 e^{rt_0} e^{r\tau} \iff 1 = ae^{r\tau} \iff \frac{1}{a} = e^{r\tau} \iff \log\left(\frac{1}{a}\right) = r\tau \iff \tau = -\frac{1}{r}\log(a).$$

Hence it takes time $-\frac{1}{r}\log(a)$ after harvesting for the population to recover. This answer makes empirical sense: because $0 < a < 1$ this is a positive time, and the smaller a is (i.e. the larger the proportion of population we harvest is) the longer the time to recover is. Similarly, the higher the birth rate is, the shorter the time for the population to recover is.

Nicely observed.

(b) The same student's revised attempt, which was graded as 'Excellent/Exemplary' (E).

Figure 3. An example of student work, illustrating typical feedback given on a first attempt and how it was acted upon by the student during the revision stage; annotations and feedback given by the marker are shown.

3.2. Staff views

We discuss the benefits and challenges of the changes made to the assignment regime as perceived by the course lecturer, Nikola Popović, and one of the two graduate TAs who provided written feedback to the lecturer.

The teaching team observed that students seemed to discuss their submissions, and the feedback on those, in informal 'study groups', in line with our aim to establish mathematics as a genuinely collaborative activity. Still, the lecturer felt that attendance at workshops was lower than expected, although it still exceeded attendance levels of similar courses in the School of Mathematics. That was unexpected, since homework assignment problems were taken verbatim from the problem sheet for that week.

Furthermore, and despite explicit guidance, students did not seem to engage with the problem sheets prior to the workshops or focus on the problems that were likely to be assigned for submission. The TA, who assisted with workshops, believed that students who engaged with them did benefit, as workshops provided a space for students to collaborate with, and learn from, their peers. They also noticed that students' mathematical writing improved over time, as did the overall quality of student submissions. The lecturer observed that students who regularly attended office hours to engage with feedback provided by markers had the best overall outcome, which seems to reiterate again a need for fostering a culture of engagement with feedback and learning opportunities. The TA felt that student workload had increased to an extent that some of the class had struggled to engage with coursework, particularly towards the end of the semester when there was an accumulation of deadlines from other courses. That struggle was evident from the decreased number of submissions for the final homework assignment. However, it is a common occurrence for coursework submission rates to decrease over the semester; this course was not unusual in that regard.

There were challenges for the teaching team, notably with regard to reaching a shared understanding of expected standards in the course. An initial calibration meeting after the first assignment deadline, where markers discussed submitted work, was identified as being particularly helpful for setting expectations of markers. The significant number of resubmissions throughout the semester resulted in increased workload for the teaching team to some extent, too. Our expectation was that the number of resubmissions would decline with time as students adjusted to expected standards and improved their initial submissions. However, that did not happen—the number of students submitting a second attempt was consistently above 80%, and only fell significantly at the final submission deadline.

Markers' attitudes changed, as well, as did the nature of the feedback they provided. Knowing that students would directly engage with their feedback resulted in a feeling of increased responsibility to provide feedback that would result in an improved second attempt. Markers had to be careful not to overlook serious errors, as those then could not, for reasons of fairness, be held against students on resubmission even if they substantially affected the standard of their work. On the other hand, markers appreciated that they did not have to focus on trivial errors, which could simply be remarked upon and left for students to correct.

Through informal conversations in class, the lecturer found that students seemed generally open to the assignment regime, and that they understood its intended educational aims. They particularly seemed to appreciate the opportunity to resubmit an assignment after revision, mirroring the scientific process in mathematical research.

On the whole, and in spite of the scaffolding and reassurance that were provided by the teaching team throughout the semester, students experienced the approach as unfamiliar and daunting at times. Moreover, the TA felt that the regime gave an advantage to students who were more sociable with their peers, as these students were more comfortable engaging in group discussions and seeking advice from classmates on how to approach problems. However, that could also be viewed as a benefit, since one of the aims of the new assignment regime was to encourage greater collaboration between students.

Finally, and anecdotally, the students present at the final (revision) lecture for the course seemed to recognise that the assignment regime was beneficial to their learning. The class was asked to respond to two questions with a show of hands. Only one student out of approximately 25 indicated that they liked the assignment regime; however, more than half of them agreed that it had benefitted their learning in the course.

3.3. Student views

Student views were gathered through an end-of-course feedback questionnaire and through semi-structured online interviews with two students after the end of the course. The interviews were based on a prepared protocol that included open and closed questions. We obtained advance ethical approval to use students' responses, submissions, and marks in accordance with institutional procedures.

The end-of-course feedback questionnaire received 8 responses from the cohort of 75 students. Every student giving feedback expressed negative views towards the EMRN Rubric, though these were perhaps the students who were most motivated to complete the questionnaire. Students seemed most unsettled by the way letter grades in the EMRN Rubric had been mapped to numerical marks and felt that they were receiving unduly low marks for partially correct work.

> The maths in my work has been correct but I may not have got the complete solution so I'm rewarded with a fail.

> Its very demotivating to put work into a homework and receive 20% when you know that a lot more than 20% is correct.

These views support our earlier observation that our students have come to expect credit for work which is not complete. Another common view was that the EMRN Rubric did not sufficiently reward the effort that students had invested into assignments.

> The strategy of the assignment is just so bad. We write this in effort and what did we get eventually 0.5/2.5 that's rediculous and we dont have the solution

> No matter how much work I spend on the assessment, I still get the same mark for every submission.

These views suggest a mismatch in the perceived purpose of homework assignments between students and the teaching team. The assessment structure was intended to encourage and reward the production of holistically correct and well-explained mathematics. However, some students felt that their efforts ought to translate into higher marks, even when they accepted that their solutions were not correct overall. The above comments also indicate that students were dissatisfied with the mapping of grades in the EMRN Rubric to numerical marks, which may simply be due to the novelty of the assignment regime within our institution, or to mathematics students in the United Kingdom being accustomed to scoring highly on assessments (Simpson, 2016).

Around nine months after the end of the course, the lecturer invited students to reflect with him about their experiences. Two of these conversations took place. The students were asked to share their initial impressions of the course, and of the assignment regime in particular. Both students were, at least initially, positive about being able to resubmit work after receiving feedback.

> I was originally quite excited about the prospect of the homework being marked and then you can do it again and then hand it back in. (Student A)

While Student A went on to find the assignment regime 'super demoralising and super stressful', the other (Student B) 'felt like it all worked fine in practice' after they became accustomed to it.

Resonating with a theme emerging from the end-of-course questionnaire, both students commented on the mapping of letter grades from the EMRN Rubric to numerical marks. Student A highlighted that it 'felt so horrible' to receive a mark of 20% for a piece of work that was 'to an extent good'. Student B mentioned the jump in numerical mark from 'Revision Needed' (R; 20%) to 'Meets Expectations' (M; 70%) as being 'quite harsh' and 'very different to the regular university grading scheme where you might do very average and get something like 50.' However, as the semester progressed, they felt 'it did feel a lot more normal':

> [. . .] in aggregate, as you start averaging out over the assignments, it did feel a lot better.

Both students felt that the assignment regime had encouraged them to collaborate more closely with their peers on digesting feedback and helping each other to improve.

> We would kind of compare and make sure that we were trying to get kind of higher marks. (Student A)

When asked to identify drawbacks of the assignment regime, both students identified issues with workload, turnaround times, and submission deadlines.

Both students were receptive to the idea of the regime being used in other courses.

> Yeah, I mean, I did think it was quite a good idea with the resubmissions. So I would I would be happy to do another course that also use this marking scheme. (Student B)

> I feel like if we introduced it at a lower level [. . .] I think it would be extremely valuable. (Student A)

While Student B expressed contentment with the implementation in this course, Student A felt that it would be particularly helpful to introduce specifications grading earlier in the degree programme so students could become accustomed to it.

3.4. Concluding remarks

Overall, the teaching team felt that the introduction of specifications grading to the course largely met its aims. Students generally engaged with the feedback provided, applied it to their work, and improved their submissions as a result. The team also felt that students demonstrated a better understanding of the course material than in the past, and that they worked more collaboratively with their peers. Knowing that feedback would be acted upon directly by students made the teaching team more mindful of the feedback they were providing. However, having to mark nearly every submission twice also significantly increased markers' workload.

Students seemed to accept the aims of the assignment regime; they particularly appreciated the opportunity to resubmit assignments after receiving feedback. However, many felt that the mapping of letter grades to numerical marks was harsh and that their efforts were not being rewarded as they had expected. Students also felt that resubmission increased their workload for the course. These issues could be addressed by modifying the mapping of grades to marks to be more generous, reducing the number of assigned problems to give students more time to attend to feedback, or introducing further opportunities for resubmission. In particular, one could require students to achieve a specified number of 'M' and 'E' grades within the EMRN Rubric in order to receive a corresponding grade in the University of Edinburgh CMS, rather than mapping individual grades in the rubric to numerical marks.

Some of the perceived issues may be due to the novelty of the approach. Within the course, these could be mitigated by sharing (anonymised) sample work from previous years to exemplify expected standards, making available performance statistics from previous years to reassure students that improvement is possible, and giving students more time to engage with assigned problems. These are changes the course lecturer is planning to implement in the course for the academic year 2024-25. Another obvious way to reduce the novelty factor would be to use specifications grading more frequently, particularly towards the beginning of degree programmes.

References

Clark, D., & Talbert, R. (2023). *Grading for growth: A guide to alternative grading practices that promote authentic learning and student engagement in higher education* [OCLC: 1390712971]. Routledge.

Iannone, P., & 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*, *41*(1), 22–31. https://doi.org/10.1093/teamat/hrab007

King, A. (1993). From sage on the stage to guide on the side. *College Teaching*, *41*(1), 30–35. https://doi.org/10.1080/87567555.1993.9926781

Nilson, L. B. (2014). *Specifications grading: Restoring rigor, motivating students, and saving faculty time* (1st ed.). Routledge. https://doi.org/10.4324/9781003447061

Simpson, A. (2016). Assessment and its outcomes: The influence of disciplines and institutions. *Assessment & Evaluation in Higher Education*, *41*(6), 917–937. https://doi.org/10.1080/02602938.2015.1052369

Stutzman, R. Y., & Race, K. H. (2004). EMRF: Everyday rubric grading. *The Mathematics Teacher*, *97*(1), 34–39. https://doi.org/10.5951/MT.97.1.0034

Talbert, R. (n.d.). *The EMRN rubric* [Robert Talbert's blog]. Retrieved July 24, 2024, from https://rtalbert.org/emrn/

Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.