Qualitative grading in mathematics: how engaging students with grading criteria can enhance learning and meaningful feedback

Helen Martin and Jean-Baptiste Gramain

University of Aberdeen

Traditionally, assessments in mathematics are graded numerically, reinforcing beliefs many students and teachers have that the 'right' answer should receive full marks irrespective of its actual quality. Merely translating percentages into alphanumeric grades may accentuate this disconnection between quality of a response and grade, misguiding efforts of our students. While there's research on redesigning assessments and enhancing feedback in higher education, there's less attention to qualitative grading in science disciplines like mathematics. This project investigates whether qualitative grading positively affects students' 'locus of attention' by engaging with co-creation of criteria, focussing on essential features and qualities of mathematical solutions. Initial findings indicate qualitative grading can indeed be implemented in mathematics and may even be more effective and efficient than traditional marking; co-creation of criteria can stimulate longer-term reflection about qualities of mathematical work; and attention shifts from "where did I lose the mark?" to "how can I improve my response?".

Keywords: qualitative grading; assessment criteria; higher education; mathematical quality

Introduction and background

This is a re-collection of our presentation at the BSRLM conference in Glasgow in June 2025 where we wanted to create conversations around assessment in higher education mathematics. As such, this paper takes a narrative approach to sharing the initial work on this ongoing project. Our collaboration began in September 2024 during one of those recurring meetings in universities about recruitment – in this case how to encourage more secondary students to study mathematics at university and graduate students to become secondary mathematics teachers, however that discussion is for another time. The conversations soon turned to talking about assessment in mathematics, teaching mathematics, the intuitive pedagogical decisions we all make in the moment leading to many discussions about what *doing mathematics* might mean to our students, to us.

These regular musings echo Mason's (2002) 'discipline of noticing', taking time to think back on decisions made, surfacing patterns in the way we design tasks, the rationale that is often tacit and looking ahead to future plans for two Linear Algebra courses informed by our different experiences as teachers of mathematics – JB as a reader in mathematics and head of department and Helen as a mathematics teacher educator.

We spend a lot of time thinking about assessment and feedback not only as an integral part of our teaching but also because the grades, and ultimately the degrees we award students should be an accurate reflection of the *quality* of their work. And

therein lies the crux of the 'problem', how can we ensure our evaluative judgements are an accurate reflection of *quality* if we do not openly debate what we mean by quality in mathematics?

> Mathematics is the art of explanation. If you deny students the opportunity to engage in this activity – to pose their own problems, to make their own conjectures and discoveries, to be wrong, to be creatively frustrated, to have an inspiration, and to cobble together their own explanations and proofs – you deny them mathematics itself. (Lockhart, 2009, p.29)

The tacit assumption that mathematics must be assessed through exams and quantitatively graded is a habit, a cultural norm which often leads to students asking, 'where did I lose the mark?' To change this habit, we need to move away from percentages to what we mean by *quality*. What are the qualities of a good mathematician or a good piece of mathematics? What does doing mathematics mean? Can we shift the students' thinking from gained/lost marks to thinking about the mathematics as the art of explanation with the opportunity to work 'like mathematicians' (Oates et al., 2016).

Decompressing assessment in mathematics

Traditionally, assessments in mathematics at all levels are graded numerically (Iannone & Simpson, 2015). This reinforces the belief many students and teachers have that a mathematical answer is simply right or wrong, and that the 'right' answer should receive full marks irrespective of its actual quality. In turn, this can lead to feedback which is poorly understood or unactionable. Merely translating percentages into an alphanumeric grade may accentuate this disconnection between the quality of a response and its grade, lead to grade inflation, and misguide the efforts of our students.

There is a large body of work on how to redesign assessment and enhance feedback in higher education in general (e.g. Boud & Falchikov, 2007; Bloxham & Boyd, 2007; Hounsell, 2007, 2024; Sadler, 1989, 2009) however, there is less attention to qualitative grading within specific disciplines such as 'pure' sciences. This project investigates this in the context of mathematics, asking whether the use of a non-traditional grading system could positively affect the 'locus of attention' of students, and thus their learning and motivation (Kluger & DeNisi, 1996), while avoiding negative impacts where 'criteria compliance' replaces 'learning' (Torrance, 2007). The underlying premise is that through engaging students in the co-creation of grading criteria, assessment becomes an essential and integrated part of the learning process – formative assessment. We are mindful of the work of Crossouard and Pryor (2012) who used formative assessment as a space to consider the 'entanglement of theory *with* practice' using Barad's (2007) notion of boundary-drawing practices to disrupt dominant discourses and practices where formative assessment could become more 'educational' where:

> …opportunities to focus on the contingencies and politics of our meaning-making were sometimes taken up more openly and dialogically with students. (Crossouard & Pryor, 2012, p.251)

In this case study, classes studying a fourth-year course on Galois Theory in 2022/23 and a second-year course on Linear Algebra in 2023/24 and 2024/25 took part in a series of activities focusing on making sense of qualitative grading and grading criteria.

Design phases- a pedagogical mise-en-abîme

Even though undergraduate students are used to receiving grades on an alpha-numerical scale, they tend to have little or no experience of qualitative grading in mathematics. In order for them to benefit more from this approach, a number of activities have been designed and implemented with students on two undergraduate courses, a fourth-year course on Galois Theory (ca. 20 students) and a second-year course on Linear Algebra (ca. 50 students). The University of Aberdeen (UoA, 2014) uses alpha-numeric grades on the Common Grading Scale (CGS) with two versions: one used for courses which are 'predominantly essay-based', and one for those which are 'predominantly numerical-based'. It is the latter which was used in this work. The development of qualitative grading came in chronological phases, each building on and extending previous adaptations (figure 1).

Phase 1 involved students studying Galois Theory (Gramain, 2023) during academic year AY22/23. After discussing the university-level CGS grade band descriptors (UoA, 2014) used for all disciplines and their relevance to the course, students were presented with five example solutions to a simple exercise provisionally graded as A to E to open up conversations around what differentiates between grade bands. The distinction between bands A (Excellent) and B (Very Good) is of particular interest, because it can often encapsulate the difference between a solution which contains all the required arguments (and is therefore 'right' but only 'Very Good') and one which presents all these arguments with the high degree of precision, rigour and clarity one would expect from, for example, a mathematics textbook.

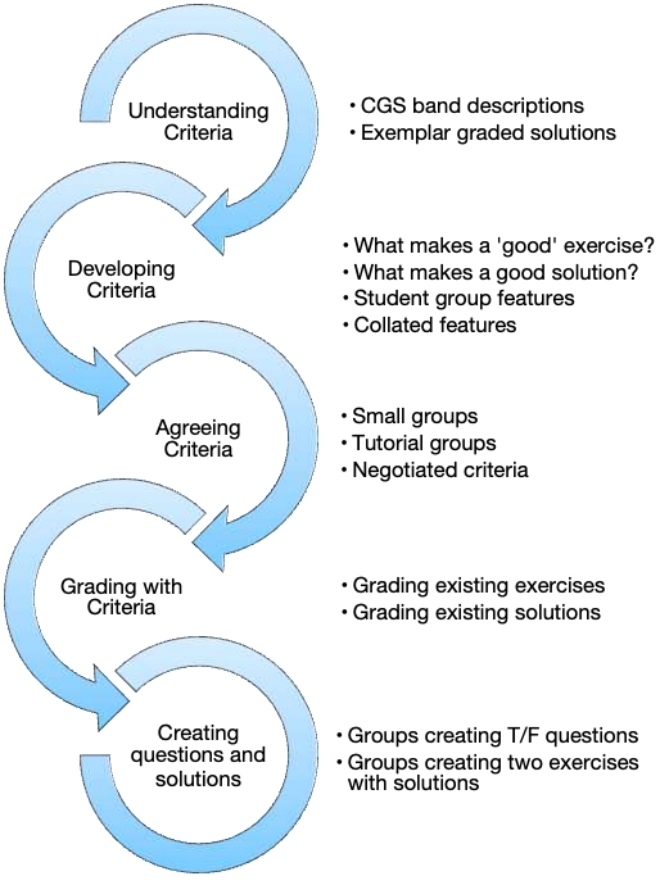

Figure 1: Assessment in Higher Education Mathematics design model.

At this time there were two homework assessments and a final closed-book exam for the course, all to be qualitatively graded directly using the CGS. Before submitting their first assessment, students were given the option to submit sample solutions to tutorial questions, to be graded using the CGS by the course coordinator.

Phase 2 AY23/24, the ideas from Galois Theory were built into the second year Linear Algebra course. After a session on grade descriptors similar to that described above, students were invited to discuss, in small groups, the qualities which make a good mathematics exercise, and those which make a good solution. Coming back as a class to exchange their ideas, students thus co-created their own criteria specific to the topic. They were then given the chance to use these to evaluate existing tutorial questions and solutions. All these activities took place during the tutorial sessions, each tutorial taking place three times for small groups of students rather than the whole class.

As part of their summative assessment for the course, students were then asked to produce, in small groups, first a set of ten True/False questions, and later two exercises complete with full solutions. These two submissions contributed 20% of the final grade for the course and were graded using the negotiated criteria. The remaining 80% consisted of a closed-book exam, also graded using the CGS.

Phase 3 AY24/25, further extended the ideas and activities of Phase 2. Rather than using the tutorial sessions, dedicated sessions were timetabled for the whole class to take part in the activities. In addition, the final closed-book exam was removed, the weight associated to the two group submissions described above was increased to 25% each, and a third, individual assessment was introduced, also worth 25%, with students having to produce three exercises, complete with solutions.

Emerging thoughts…

The iterative task design and continuing reflexive conversations allow thoughtful adaptations including extending understanding predetermined to co-constructing negotiated criteria and from using exemplar solutions with grades to using existing exercises and solutions without predetermined grades to minimise ‘compliance’ and further open up debates around what differentiates very good from excellent. Some adaptations are ongoing such as the work around students creating their own questions (Swan, 2005, 2006). Given the shift to *qualitative* grading, there is also a need to reconsider the initial assumption to use the ‘numerical-based’ rather than ‘essay-based’ descriptors (UoA, 2014) for example, replacing an ability to “express arguments with a high *level of precision*” with “original insight and power of analysis in which arguments are *cogent and well supported* in almost all respects or demonstrate creativity or originality” or “ability to present well-structured and *persuasive* argument” [our italics].

We are in the process of analysing the data consisting of naturally occurring documentary traces of these activities, alongside course feedback forms and external examiner reports and so this section is at the moment emerging thoughts. Although there are already unexpected criteria being suggested by students such as a good exercise should be “thought provoking and stimulate inquisitiveness”, “challenging, forces us to think and allows creativity”. Interestingly the qualities of a ‘good’ solution included presenting “common mistakes”, “multiple ways to solve” and “whether the solution is unique”. Even though these are tentative, there is an indication that the use of co-constructed qualitative grading (a) can indeed be implemented in mathematics and may even be more effective and efficient than

traditional marking, (b) can stimulate students' longer-term reflection about the qualities required in any mathematical work and finally (c) can shift the 'locus of attention' from "where did I lose the mark" to "how can I improve my response".

References

Barad, K. M. (2007). *Meeting the universe halfway.* Duke University Press.

Bloxham, S., & Boyd, P. (2007). *Developing effective assessment in higher education: a practical guide*. Open University Press.

Boud, D., & Falchikov, N. (2007). *Rethinking assessment in higher education: learning for the longer term*. Routledge.

Crossouard, B., & Pryor, J. (2012). How theory matters: Formative assessment theory and practices and their different relations to education. *Studies in Philosophy and Education, 31*(3), 251–263. https://doi.org/10.1007/s11217-012-9296-5

Gramain, J-B. (2023, March 3). The house cup: promoting qualitative grading in mathematics. *Times Education Higher Campus*. https://www.timeshighereducation.com/campus/house-cup-promoting-qualitative-grading-mathematics

Hounsell, D. (2007). Towards more sustainable feedback to students. In B. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education: learning for the longer term* (pp.101–113). Routledge.

Hounsell, D. (2024). Feedback interchange in small-group discussion: an interpretive review of the literature. *Practitioner Research in Higher Education Online First,* 161–174.

Iannone, P., & Simpson, A. (2015). Students' preferences in undergraduate mathematics assessment. *Studies in Higher Education, 40*(6), 1046–1067.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.

Lockhart, P. (2009). *A mathematician's lament: how school cheats us out of our most fascinating and imaginative art form*. Bellevue Literary Press.

Mason, J. (2002). *Researching your own practice: the discipline of noticing.* Routledge.

Oates, G., Paterson, J., Reilly, I., & Woods, G. (2016). Seeing things from others' points of view: collaboration in undergraduate mathematics. *Problems, Resources, and Issues in Mathematics Undergraduate Studies, 26*(3), 206–228. https://doi.org/10.1080/10511970.2015.1094683

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119-144.

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment and Evaluation in Higher Education, 34*(2), 159–179.

Swan, M. (2005). *Improving learning in mathematics: Challenges & strategies*. DfES.

Swan, M. (2006). *Collaborative learning in mathematics: a challenge to our beliefs and practices*. National Institute of Adult Continuing Education.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy and Practice, 14*(3), 281–294.

University of Aberdeen (2014). *Common grading scale and band descriptors.* https://www.abdn.ac.uk/staffnet/education/assessment/cgs-scale/common-grading-scale-cgs/