The University of Edinburgh

Principal's Teaching Award Scheme

PTAS Application Proposal Cover Page

Names and titles of all team members:

Principal Applicant: Paul Anderson
School: Informatics
Building / Room No: Informatics Forum 1.24

Tel: 0131 651 3141
Email: dcspaul@ed.ac.uk

Co-applicant: Susan Rhind
School: Veterinary Medicine

Tel: 0131 650 6219
Email: Susan.Rhind@ed.ac.uk

Project Title: Supporting Criteria-Based Marking

Abstract (up to 250 words):

Criteria-based marking is capable of providing more reliable assessment and more meaningful feedback to the students. Using finer-grained criteria provides more specific feedback, and helps to mitigate the variation which inevitably occurs with multiple markers and large classes. However composing these criteria in a meaningful way is not always straightforward, especially when a fine-grained numerical result is required - as for the Common Marking Scheme.

For a number of years, we have been experimenting in Informatics with software to compose larger numbers of fine grained criteria. This has the potential to produce a fine-grained, interpolated result, and to identify specific strengths and weaknesses as feedback to the student. It also allows the criteria to be easily adjusted to ensure that the result represents a meaningful assessment of the appropriate objectives.

We have recently been working to create a version of this experimental software which is usable by other courses. The aim of this project is to evaluate this software, and the criteria-based approach in general, in a different context - specifically vet medicine. This work has the potential for impact across multiple contexts and disciplines.

Institute for Academic Development

Background

As Sharron Ogle asks in the Teaching Matters blog [6], "What does a numerical mark even mean, in an environment where there is no single correct answer and there are very few facts?". This is a difficult question even in a subject such as Informatics which might be assumed to have more clear-cut answers than many other disciplines. However, our courses are required to generate a final numerical mark, currently on the University's 100-point "Common Marking Scheme".

Criteria-based marking [5, 8] is an increasingly popular alternative to "additive" or "percentage" grading. This involves identifying a set of criteria corresponding to the learning outcomes of the course and judging how well each criteria has been met. Aligning judgements to such criteria provides a much richer environment in which to embed meaningful and constructive feedback to students. Accepting that we need to then transform these judgements into a final numerical mark, this can be done in several ways: A "rate-weight-sum" process is not uncommon, but this is "compensatory" and is claimed not to have a "sound theoretical rationale" [9]. Composition of the criteria using conjunctive/disjunctive rules is probably the most common approach, but this is not without problems either: in practice, "tacit knowledge and implicit norm referencing are often used to interpret the criteria and justify a more holistic decision" [2].

Whatever the marking approach, there is also the common problem of consistency across multiple markers, and the need to explain the reasoning behind the resulting marks to the student. These become more problematic as the student numbers increase.

For a number of years, we have been using a criteria-based approach to assessing a masters-level programming course [1], and we have been gravitating towards larger numbers of criteria with simpler judgements. Especially if these are spread across multiple markers, this can help to minimise variations due to differing interpretations of individual criteria. And the results of this can be used to automatically provide more detailed feedback to the students on their particular strengths and weaknesses.

For example, a "code readability" criteria may be represented by a number of questions, marked on a simple 4-point lickert scale (which we think of as: *no*, *not-really*, *sort-of* and *yes*):

- Is the code properly indented?
- Do the large-scope variables have meaningful names?
- Are there adequate comments?
- Are there too many comments?
- Is there redundant commented-out code?
- Are there methods which are too large?
- …

However, both Bloxham [3] and Sadler are wary of this direction, noting that "the sets of objectives tend to become atomistic and unmanageable". But we have been developing software [1] to assist in the management of the criteria and we are hopeful that this might help to alleviate some of the difficulties. In particular, the software allows us to easily combine larger numbers of rules in ways which are intended

---

[1] https://www.inf.ed.ac.uk/teaching/courses/ippo

to reflect higher-level criteria (such as the above "readability"). And it allows us to retrospectively adjust the rules until they provide a more authentic, and explicit representation of a desired holistic assessment.

PMark provides a simple language to express logical decision rules for the various assessment grades. It will also interpolate automatically between those grades if a finer-grained result such as a percentage is required. This is in contrast to systems such as [4] which use fuzzy logic to determine the grade boundaries: the boundary decisions in PMark are always explainable in terms of the basics rules, and we are hoping to extend the system to provide these explanations automatically.

The Proposed Project

In veterinary medicine (and medicine), recent assessment trends are moving towards a more programmatic assessment approach, recognizing that the whole is more than the sum of the parts. Furthermore, our current curriculum review in veterinary medicine has a key focus on changing assessment to align more with criteria based marking hence there is an ideal window of opportunity to perform empirical research that will have a direct impact on curriculum and assessment development. This project will provide the opportunity to compare the development of criteria-based marking schema in different disciplines and explore from both the staff and student perspective experiences of working with the tool and engaging with feedback facilitated by the tool respectively.

Aims & Objectives: We would like to:

- Evaluate the suitability of PMark for courses outside of Informatics, specifically Veterinary Medicine.
- Improve the software as necessary in response to this evaluation.
- Explore how the use of the criteria can be used to support the delevelopment of assessment literacy.
- Experiment with, and evaluate the automatic generation of feedback.
- Disseminate software, results and documentation.

Methodology: We will develop schema for existing courses in both Informatics and Veterinary Medicine, and use these as part of the assessment. We will evaluate the results using qualitative techniques such as focus groups, involving both staff and students. We would also hope to compare marks from existing assessment schemes (such as those using additive criteria) with those generated by a PMark scheme.

In addition to the technical aspects of project, we will also use the new schema in assessment literacy interventions in each course as part of the project. For example, tutorials will be designed to allow students to work with the criteria as they assess the quality of other pieces of work against them in a similar way to previous studies we have published [7].

Dissemination & Impact: The results of this study will have implications not only for the disciplines under study but also for any subject areas looking to move towards a more criteria based approach to assessment and feedback. Through constructive alignment with course learning objectives, this work is also highly relevant in the context of curriculum review and any potential future review of the University Common Marking scheme. We will ensure impact is optimised by disseminating through the routes outlined below.

We will also make the software freely available under an open-source license.

Dissemination

Audience	Event	Detail
School/ University	School Learning and Teaching Committees and local workshops.	Both students and staff at the school will be kept up to date with the findings and developments of the study and offered the opportunity to input to the outcome.
	Relevant IAD or other University events/ committees.	E.g. University Learning and teaching Conference 2021.
National/ International	VetEd symposium 2021 (Nottingham).	Annual conference attracting 150+ delegates from veterinary education community national and international attendance.
	Assessment and Evaluation in Higher Education Journal article.	We aspire to publish in this generic HE journal in recognition of the interdisciplinary approach to the study design.

Timetable

June - October 2020	Develop schema in collaboration with discipline specific staff.
October - December 2020	Embed in relevant courses in informatics and veterinary medicine. Use schema as part of assessment literacy interventions during the roll-out of the course.
January - April 2021	Qualitative evaluation with both staff and students.

Budget

Research Assistance (approx 25 days)	£3827	To employ one or more postgraduate and/or undergraduate students to help manage the project, develop the tool and analyse focus group data.
Audio-transcription	£130	In order to subject interview data to robust thematic analysis, audio-transcription is required. Rate based on transcription of 120 minutes of recordings with a two week turnaround.
Focus group refreshments for staff and student focus groups in semester 2	£100	
Conference fees x 2	£600	Fees for dissemination at relevant discipline specific and/ or generic higher education conferences.
TOTAL	£4,657	

References

[1] P. Anderson. Pmark: A tool to support criteria-based marking. https://homepages.inf.ed.ac.uk/dcspaul/homepage/live/work/pmark.html.

[2] S. Bloxham, P. Boyd, and S. Orr. Mark my words: the role of assessment criteria in UK higher

education grading practices. *Studies in Higher Education*, 36(6):655–670, 2011. https://doi.org/10.1080/03075071003777716.

[3] S. Bloxham, B. den Outer, J. Hudson, and M. Price. Let's stop the pretence of consistent marking: exploring the multiple limitations of assessment criteria. *Assessment & Evaluation in Higher Education*, 41(3):466–481, 2016. https://doi.org/10.1080/02602938.2015.1024607.

[4] J. A. Montero, R. M. Alsina, J. A. Morán, and M. Cid. Fuzzy logic system for students' evaluation. In J. Cabestany, A. Prieto, and F. Sandoval, editors, *Computational Intelligence and Bioinspired Systems*, pages 1246–1253, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg. https://link.springer.com/chapter/10.1007%2F11494669_153.

[5] B. M. Moskal. Scoring rubrics: What, when and how? In *Practical Assessment, Research, and Evaluation*, volume 7. 2000. https://scholarworks.umass.edu/pare/vol7/iss1/3.

[6] S. Ogle. Teaching matters blog: Assessment: what is it for? https://www.teaching-matters-blog.ed.ac.uk/assessment-what-is-it-for/ [cited 19th February 2020].

[7] S. M. Rhind and J. Paterson. Assessment literacy: Definition, implementation, and implications. *Journal of Veterinary Medical Education*, 42(1):28–35, 2015. PMID: 25547906. https://doi.org/10.3138/jvme.0714-067R1.

[8] D. R. Sadler. Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30(2):175–194, 2005. https://doi.org/10.1080/0260293042000264262.

[9] D. R. Sadler. Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34(2):159–179, 2009. https://doi.org/10.1080/02602930801956059.

Applicant Backgrounds

Paul Anderson: is a Senior Research Fellow with the School of Informatics[2]. Paul has been teaching programming for nearly 40 years and is currently responsible for the IPPO MSc programming course. He is organiser for the Informatics first-year, and a tutor on first-year programming courses. Aside from his research interests in large-scale system configuration, Paul has an ongoing interest in education – he has a PGCE and has been involved in several PTAS projects.

Susan Rhind: BVMS, PhD, FRCPath, PFHEA, MRCVS, is Director of Veterinary Teaching and Chair of Veterinary Medical Education at the Royal (Dick) School of Veterinary Studies. Her research interests include assessment and feedback, e-learning, and curriculum development.

---

[2]homepages.inf.ed.ac.uk/dcspaul