

Serial number 0136

Title SAFE: Self-Assessment and Feedback for Education

Submitter Dr. Ross Davies, Dr. Marius Miknis

Introduction

SAFE is designed to inspire good practice and consistency in all areas related to assessment. One of those areas is Self-Assessment. Self-assessment potentially helps improve a student's understanding of their own grade. The implementation of projects such as SAFE is designed to help improve a student's understanding and analysis the patterns of lecturer assessment to help improve their understanding of their students. Students are asked to assess lecturers; SAFE provides 360-degree feedback to help both parties improve with the overall goal of increasing students understanding, experience and ultimately their classification.

Background

SAFE is based upon the principles of rubrics. Rubrics are two-dimensional matrices. The rows represent the different criteria used for marking (e.g. *Understanding of the material, Writing style, Demonstration ...*) while the columns contain the evaluation as standardized in the UK (*Fail, Narrow Fail, Pass, 2.2, 2.1, First*). Each individual cell is made of one or more tokens describing in detail what a grade represents given a specific criterion. For example, a *Pass* in the *Writing style* section could be explained by the token: "Writing style is appropriate".

Rubrics have numerous benefits for both the lecturers and students. These benefits translate across subject areas ultimately improving the grading process for both parties. Rubrics have been shown to increase the grading accuracy of the assessors [1] [2]. In comparison to pen-and-paper based methods electronic based methods can improve assessment times by up to 200% [3]. Rubrics also provide immediate feedback showing the students a breakdown of exactly what they achieved [4]. Different tools have been created to support and implement these findings [5] [6] but also to provide a qualitative extensive feedback [7].

Students also benefit from Rubrics as self-assessment allows a student to evaluate their own work against the same criteria as the official assessor. Comparison can be made against the official assessor and the student. Evaluating the success of self-assessment typically assumes that the official assessors grade is the gold standard [8]. Jonsson and Svingsby [2] argue that differences between instructor and student judgments might well be attributed to the students' weaker understanding of the criteria used and not to their performance as such.

Method

SAFE allows an assessor to create an assessment brief and associated rubric. Once the submission date is passed the self-assessment is made available (if an assessor wishes to do so they can release the self-assessment earlier). It is recommended that the students self-assess first. However, the official assessor cannot see the students grade until after they complete their assessment. Inconsistencies are then made available to the assessor with a prompt if there is an area where the student overestimated their ability. Either for review (quality procedure) or for additional feedback.

Once the assessor has completed the assessment and no longer wishes to make any additions a feedback document is compiled for the student. This will show: the selected portions of the Rubric; assessor feedback; self-assessment review highlighting any inconsistencies on a criteria basis rather than an individual token; a breakdown of the grade showing how much they achieved in the coursework and the contribution to their final grade.

Data Analysis

Attempting to analyse the impact of self-assessment on the student's grade can be achieved in several ways. The methods chosen in this paper are: Pearson's correlation (see Table 1), Spearman's RHO based on the classification boundaries (see Table 2) and measuring the accuracy within five-percent of the assessor's grade (see Table 3). Two different samples were used with a different cohort of students in each sample. Sample 1 contains 46 participants and 2 assessments. Sample 2 contains 34 participants and 6 assessments.

Table 1: Persons Correlation Coefficient

Sample 1		Sample 2					
A1-1	A1-2	A2-1	A2-2	A2-3	A2-4	A2-5	A2-6
0.35	0.46	0.32	0.62	0.75	0.68	0.80	0.86

The correlation can be visualised using a scatter diagram where the x-axis represents the assessors grade and the y-axis represents the self-assessment.

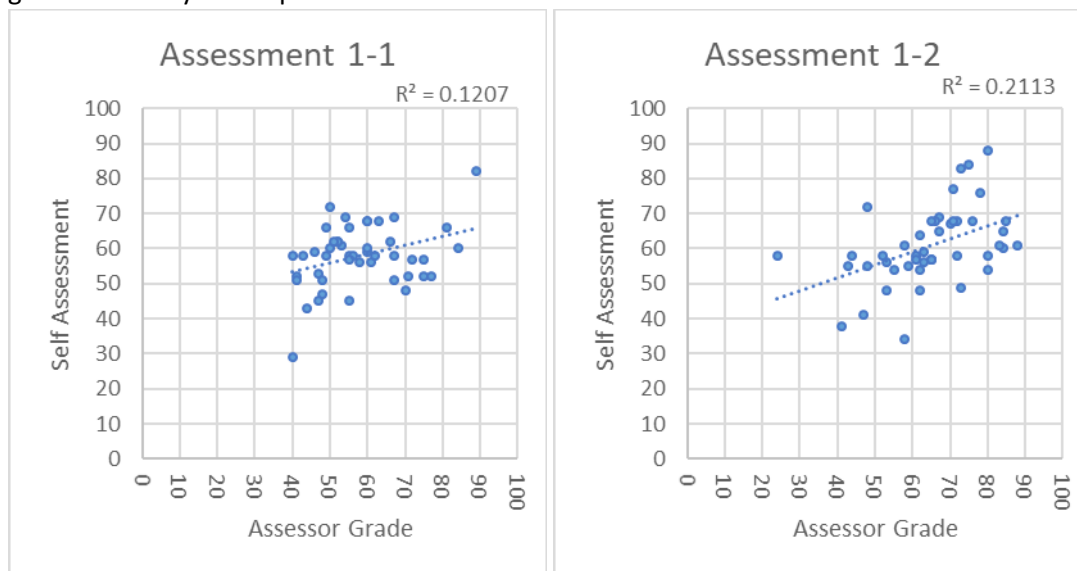


Figure 1: Sample 1 - Self-Assessment Analysis

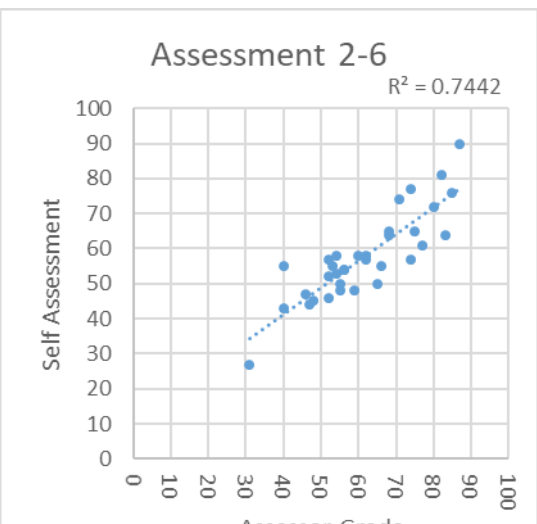
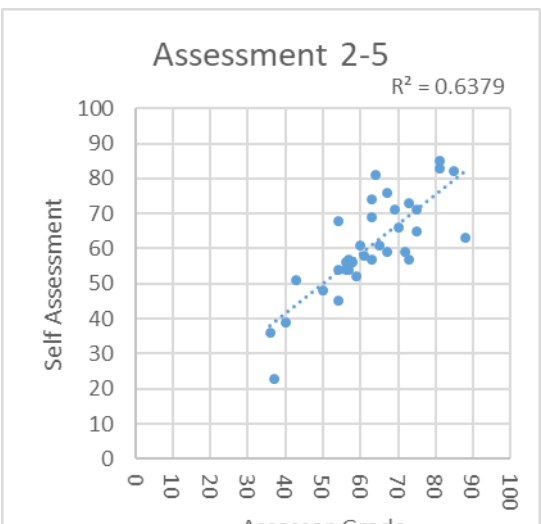
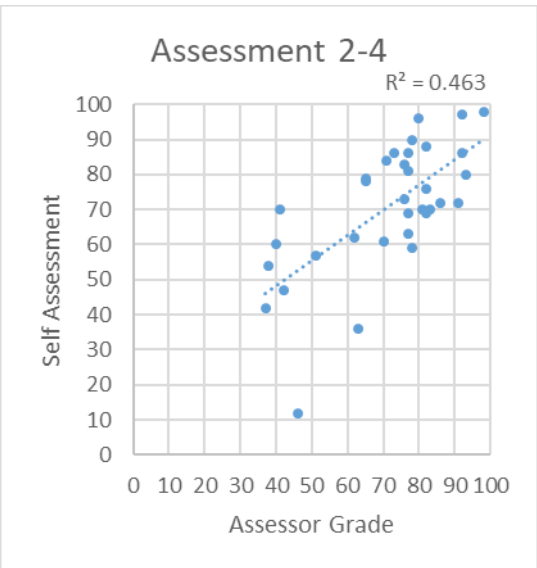
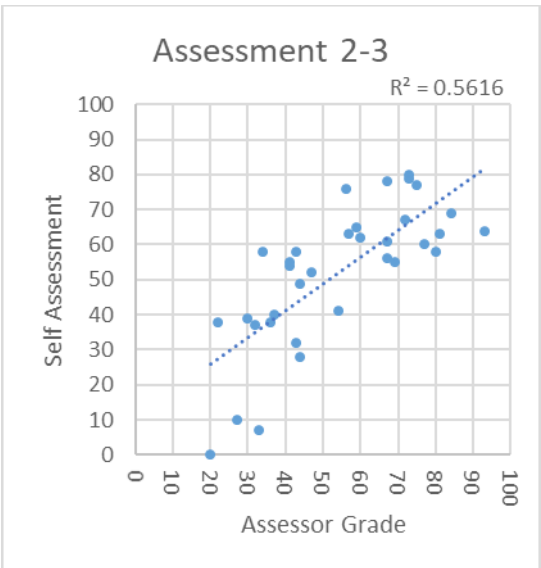
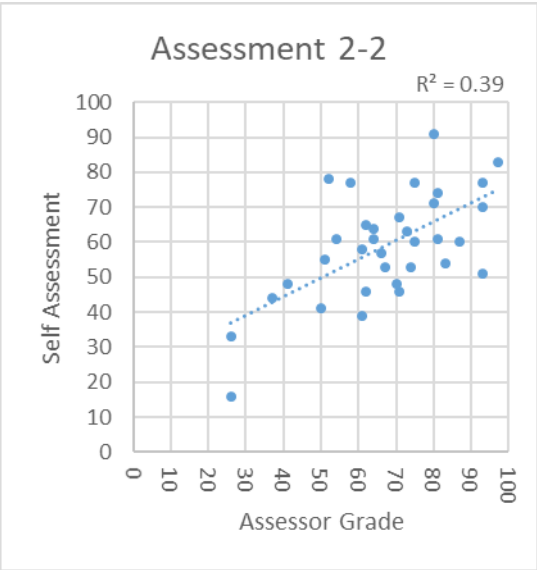
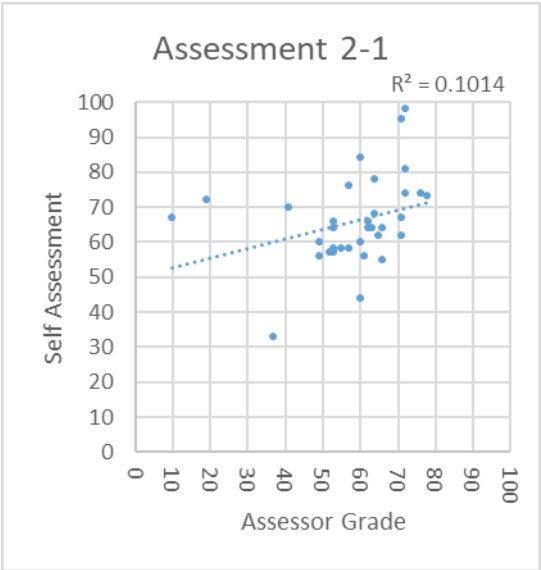


Figure 2: Sample 1 - Self-Assessment Analysis

Classification Analysis:

A significant boundary for students tends to be their classification. A Spearman's RHO analysis of the classification boundaries shows a significant improvement in the student's ability to self-assess accurately. The first sample shows the correlation value transition of 0.28 increasing to 0.53 over the two assessments. In the second sample, there is a similar transition over the first two assessments.

Table 2: Spearman's RHO Analysis based on classification (Sample 1 - 6 Assessments and Sample 2 - 2 Assessments)

Sample 1		Sample 2					
A1-1	A1-2	A2-1	A2-2	A2-3	A2-4	A2-5	A2-6
0.28	0.53	0.39	0.51	0.80	0.57	0.70	0.85

Accuracy within 5 marks

The results for sample 1 are shown below. Assessment 1 was designed to get a base line and figure out how well the students had taken in the simple topics. This meant there was a higher ability to self-assess which could be linked to the simplicity of the students. However, even in this assessment some students had trouble self-assessing with the maximum. In the second sample, the same trend can be seen although due to the small number of assessments a 6% increase would be within a margin of error. Further in class tests would be necessary with this cohort.

Table 3: Accuracy within 5%

Sample 1		Sample 2					
A1-1	A1-2	A2-1	A2-2	A2-3	A2-4	A2-5	A2-6
15	18	18	7	8	7	19	22
33%	29%	53%	21%	24%	21%	56%	65%

Conclusion & Future Work

Self-Assessment is vital to ensure students are understanding the marking scheme. As academics, we have a responsibility to ensure that the assessments we set are meaningful and the feedback is constructive. However, in most circumstances we only know the overall grade and what the student achieved but if they don't contact us to query something we rarely have knowledge of their understanding. The evidence demonstrated here shows a significant improvement in the student's ability to estimate their classification after a small number of assessments.

Bibliography

- [1] M. Bresciani, M. Oakleaf, F. Kolkhorst, C. Nebeker, J. Barlow, K. Duncan and J. Hickmott, "Examining design inter-rater reliability of a rubric measuring research quality across multiple disciplines," *Practical assessment, Research & Evaluation*, vol. 14, no. 12, pp. 1-7, 2009.
- [2] A. Jonsson and G. Svingby, "The use of scoring rubrics: reliability, validity and educational consequences," *Educational Research Review*, vol. 2, no. 2, pp. 130-144, 2007.
- [3] L. Anlgin, K. Anglin, P. Schumann and J. Kaliski, "Improving the efficiency and effectiveness of grading through the use of computer-assisted grading rubrics," *Decision sciences journal of innovative education*, vol. 6, no. 1, pp. 51-73, 2008.
- [4] D. Stevens and A. Levi, *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning*, Stylus Publishing, LLC, 2005.

- [5] T. Auvinen, "Rubric," in *International conference on computing education research*, Koli, 2011.
- [6] L. Gardner, D. Sheridan and N. Andreas, "Rubric marking "out of the box": Saving time & adding value to teaching & learning," *World conference on educational multimedia, hypermedia and telecommunications*, vol. 2011, no. 1, pp. 391-400, 2011.
- [7] T. Ahoniemi and V. Karavirta, "Analyzing the use of a rubric-based grading tool," *ACM SIGCSE*, vol. 41, no. 3, pp. 333-337, 2009.
- [8] M. Ward, L. Gruppen and G. Regehr, "Measuring self-assessment: current state of the art," *Advances in health sciences education*, vol. 7, no. 1, pp. 63-80, 2002.
- [9] K. Cho, C. Schunn and R. Wilson, "Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives," *Journal of Educational Psychology*, vol. 98, no. 4, p. 891, 2006.
- [10] K. Lasater, "Clinical judgment development: Using simulation to create an assessment rubric.," *Journal of nursing education*, vol. 46, no. 11, pp. 496-503, 2007.