1   *Correlating peer and tutor-assessment on a low-stakes engineering*

2   *assignment*

3

4   **Diogo Montalvão[1] and Thomas Baker**

5   School of Engineering and Technology, University of Hertfordshire, United Kingdom

6

Abstract

8   Peer-assessment has been a subject of great debate in recent years. The

9   way students perceive assessment and what motivates them when

10 assessing, may differ significantly from the tutor. This paper discusses a

11 study designed to correlate student's marking with the marks awarded by

12 their tutors when peer-assessing one another from in-class oral

13 presentations. A new and alternative approach to correlate results is

14 presented, which is based on the normalisation of the quantitative

15 judgements based on determined criteria. The methodology was *blind* and

16 *holistic*, as described in previous works: some guidelines were provided to

17 the students on what is considered acceptable without getting into detail

18 (*holistic* marking) and peer-assessment marks were made confidential

19 (*blind* approach). It was observed that students have a tendency to overrate

[1] **Corresponding author:** Diogo Montalvão, School of Engineering and Technology,

University of Hertfordshire, College Lane Campus, Hatfield AL10 9AB, UK. Email:

d.montalvao@herts.ac.uk

20   fellow students - especially where lower marks might be awarded. There

21   is, however, direct agreement with the tutor's marking in terms of

22   qualitative judgements, which is highlighted by the presented correlation

23   method used to adjust students marks. The presented methodology to

24   correlate marks between the students and tutor showed to be a promising

25   one. After processing the data with this simple and straightforward

26   algorithm, peer- and tutor assessment practically showed a perfect match.

27   **Key words** peer-assessment; assessment for learning; higher education

28   assessment; assessment normalisation.

29

30   Introduction

31   Orsmond (2011), in his review on assessment within the higher education context,

32   discriminates three ways of considering assessment in relationship to learning:

33   assessment *of* learning, assessment *for* learning and assessment *as* learning. The widely

34   accepted practice for assessment still is in terms of being *of* learning. Typically, students

35   are assessed through summative assessment, e.g., an end of module examination in

36   which the student's amount of learning is fully assessed by the tutor. However, Boud

37   and Falchicov (2006) consider this approach as short-termed under the context of higher

38   education, in which new material must be gradually appreciated through an iterative

39   updating process, so that life-long self-learning is stimulated after graduation.

40   The conceptual shift of assessment from "*of* learning" to "*for* learning" and then

41   to "*as* learning" can be traced back to Gipps (1994) and Earl (2003), respectively.

42   Although with different emphasis, both acknowledge the introduction of student self-

assessment, in which the student assumes an important role in the assessment and
judgment of his own work. Another frequently adopted method in which students are
involved in the appreciation and appraisal of learning is peer-assessment (Van Gennip
et al. 2010). According to Orsmond (2011), "peer-assessment shares most of the key
features of self-assessment" and "feeds self-assessment activities particularly through
the act of receiving and giving feedback." Prompt feedback is appreciated by students,
because it gives them an early opportunity to assess their understanding of the course
material, as acknowledged by Cotner et al. (2008) when introducing scratch-off
immediate feedback assessment technique (IF-AT) forms and classroom response
system (clickers) in large cohorts of students.

For peer-assessment to work effectively, some authors argue that one of the
requirements is that the student and the tutor discuss and agree – or at least clarify and
reflect on - assessment criteria (Falchikov and Boud 1989). Tsivitanidou et al. (2011),
on a study investigating secondary school students' unmediated peer assessment skills,
observed that students already have the basic skills required for the implementation of
peer-assessment, although they recognised that "reciprocal peer assessment is a rather
complex procedure that requires well developed peer assessment related skills" and that
"students need to receive explicit training". However, Norton (2004) points out that
students may become strategic learners who end up selecting what they identify as
being the key aspects to obtain marks instead of spreading their attention to the whole
span of the learning range. On the other hand, one might argue that students tend not to
perceive assessment criteria as tutors do. Most often, they are not provided with the
proper tools that would enable them to self-monitor their own learning (Hatzipanagos
and Rochon 2010) and they still have little involvement in assessment design, marking
and evaluation (Orsmond 2011).

68   Nevertheless, peer-assessment is an approach to learning that is nowadays

69   recognised as a tool capable of getting the students more involved, making them take

70   more responsibility, encouraging self-critical analysis and self-evaluation of student

71   work and fostering debate and communication (Hunter 1999 and Dochy and McDowell

72   1997). Even if a poor correlation between students self-assessed marking and the tutor

73   marking is observed, some authors (e.g., Kirby and Downs 2007) continue advocating

74   in favour of self-assessment for learning – which is strongly correlated to peer-

75   assessment according to Orsmond (2011).

76   Peer-assessment can be used as a tool to give students an opportunity to learn by

77   making them aware of the different approaches employed by other students: an aspect

78   highlighted by Van Gennip et al. (2010) by mentioning the social and interpersonal

79   bonds involved in peer-assessment, from which students, by assessing one another, are

80   actually learning from their peers. For example, Chuck and Young (2004) designed a

81   formative assessment task to improve the scientific report writing skills of university

82   students. This used a combination of peer- and self-assessment against specific criteria,

83   where students were required to submit an amended report. It was shown that those that

84   participated in this cohort-driven assessment task got better results on average compred

85   with those that would have been obtained after the first submission.

86   With the above points in mind, a peer-assessment assignment was introduced in

87   the modules of Technical Design and Geometrical Modelling I and II (1st year students)

88   of Engineering undergraduate degrees taught in *institution removed for blind review*.

89   The peer-assessment case study discussed in this paper took place during the

90   semester's last week (before examinations) and was focused on a part of the student's

91   group assignment. The students were invited to mark one another's oral presentations

92   on sheets of paper (similar to voting bulletins). These were later returned to the tutor in

93   such a way that the peer-assessment marks were kept anonymous.

94   Using the terminology from Russel et al. (2006), the followed methodology was

95   *blind* and *holistic* as opposed to *open* and *category-based*. These authors describe these

96   terms the following way: in a *blind* approach, students respond to the peer-assessment

97   task in complete isolation from one another (anonymous marking), whereas in an *open*

98   approach students discuss their thoughts and negotiate marks when peer-assessing

99   within a group. In a *holistic* marking criteria style, students are asked to evaluate their

100   peers' contribution and performance by looking at the 'big-picture'. Some guidelines

101   can be provided to the students on what is considered acceptable, but the students are

102   not asked to explicitly respond to them in any detail. On the other hand, on a *category-*

103   *based* marking scheme, students are expected to mark their peers' performance against a

104   specific set of marking criteria.

105   Results show that students are perfectly capable of making the correct

106   distinctions between one another when speaking in qualitative terms. The correlation

107   between the peer- and the tutor assessment in quantitative terms is not that different as

108   well. However, a tendency to award generous marks to weaker students is observed.

109   An alternative approach to analyse the peer-assessment marks is presented as

110   well, in an attempt to better understand how peer- and tutor assessment correlate in the

111   quantitative judgments. This is based on the normalisation of the peer-assessment marks

112   based on range and one parameter target (in this case, a value for the average was set as

113   target). For the correlation to be comparable the average of the peer-assessment marks

114   was set to the same value as the average of the marks from the tutor. The range was

115   stretched in order to consider all the possible marking range as students showed to give

116   marks on a narrower band than the tutor.

117 **Method**

118 *Data sample*

119 This study was carried out with data from *institution removed for blind review*, and

120 the assessments were part of the "Technical Drawing and Geometrical Modelling I and

121 II" modules. These are 1st year undergraduate modules taught in semesters A and B of

122 the Engineering degrees. The data analysed corresponds to four cohorts of

123 approximately 25 students each (a total of approx.. 100 students).

124 *General assessment criteria*

125 Assessment of these modules is based on a weighted combination of individual

126 coursework (both in-class and homework) and a group design project that includes both

127 an in-class presentation and a written report which is submitted at the end of the

128 semester. This is illustrated in figure 1 pie chart, in which the lower half corresponds to

129 group work and the upper half corresponds to individual work.

130
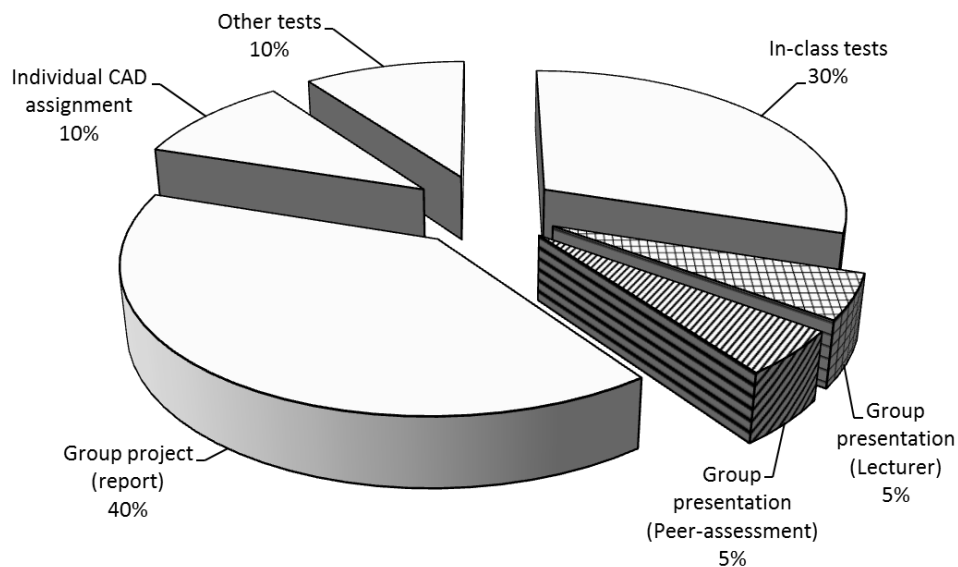
Figure 1. Marking distribution.

131

132

133   The group component of the assessment includes a report on the group project

134   worth 40% of the final marks and a group presentation worth 10%. The group oral

135   presentation is the average value between the marks given by the students when peer-

136   assessing (worth 5%) and the marks given by the lecturer (worth another 5%). Although

137   this was a group presentation, students were also assessed individually according to how

138   well they performed on the roles they played as described below (each group was

139   composed by 2~3 students, depending on the complexity of their projects).

140   *Peer-assessment marking scheme*

141   The in-class presentations ran during the semester's last week. Each group, composed of

142   2 to 3 students, prepared a presentation using both the CAD software and a slideshow.

143   Each student was given up to 5 minutes to make their presentation regarding their

144   contribution to the group work.

145   Both the teacher and the students based their assessment on the same marking

146   scale, ranging from 1 to 10 (figure 2): three levels of pass and two levels of fail were

147   considered. The zero marks were reserved for students who did not attend the

148   presentations (any student presenting would get at least a minimum of 1 mark).

149
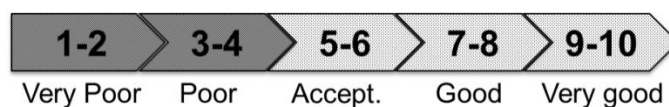
150

151   Figure 2. Marking range for the peer-assessment.

152

153   No specific marking criteria were given to the students other than to assess the

154   specific objectives defined in the coursework briefing. To elaborate a list of criteria

155 would make the task unfeasible considering time constraints and the number of oral

156 presentations each student had to assess. There also are other factors that were taken

157 into account on this decision. First, the students might feel it is the job of the lecturer

158 (and not students) to mark work, especially if a considerable amount of time is needed,

159 as pointed earlier by Falchikov (2003). Secondly, according to a study from Brown et

160 al. (1998), students tend to struggle when assessing via a discriminated marking

161 scheme, even if they already have a general perception of the quality of the work. Thus,

162 an *holistic* approach seemed more appropriate to the oral presentations, in which

163 students look at the 'big-picture'. Furthermore, this was a low-stakes part of the

164 assignment aiming at giving some guidance to the students on the progress of their work

165 from the tutor's comments.

166   For the peer-assessment to be as *unbiased* as possible, some precautions were

167 taken:

168   1. To ensure that students take it seriously and engage with the exercise, and to

169   avoid cases where students might show favoritism (e.g. by marking certain

170   students with all 10's (or 1's)), they were told that any "non-differentiating"

171   marking sheets would be rejected.

172   2. It was guaranteed that each individual marking sheet would be kept confidential

173   and would be destroyed after the results were collected. – *blind* methodology;

174   3. To make the students feel this was their free choice and not an imposition, they

175   were told they could refrain from assessing their peers and submit the marking

176   sheet blank;

177   4. To provide a sense of justice and fairness, the students could opt out from being

178   assessed by their peers. In such a case, the lecturer's assessment would be worth

179   the whole 10% of the presentation marks.

180   From the almost 100 students participating in this study, none fitted the 3rd and

181   4th categories listed above. However, there were a few whose marking sheets were

182   rejected for the reasons described in point 1 above. With respect to the 2nd point, it is

183   not possible to tell if the students feared discrimination because those sitting next to

184   them could possibly peek into their marking sheets.

185   **Results and discussion**

186   *General results*

187   A histogram plot displaying the number of times (frequency) a grade was given by both

188   the lecturer and the peers[2], is presented on figure 1. This picture includes the results for

189   both the modules together showing a trend that follows what resembles a normal

190   (Gaussian) distribution. The decision to join both the modules in a single plot is

191   justified from the fact that the conclusions that might be drawn from each of the

192   modules alone would be essentially the same. Furthermore, the statistical significance is

193   improved, since the sample so obtained increased to a size of $N=93$ elements.

194

---

[2] Marks were rounded to an integer value. For instance, grade 7 in the histogram bar chart includes all marks ranging from 6.5 to 7.4.
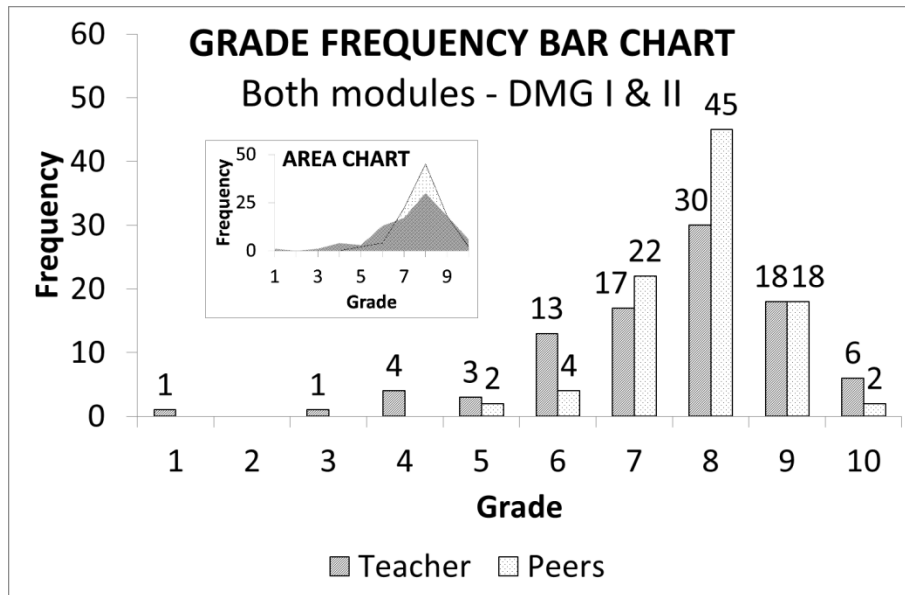
Figure 1. Histogram showing the amount of times a grade was given.

A smaller area chart is also shown in figure 1. It contains the exact same information as the bar chart, but it puts into evidence some detail that would otherwise be less marked: for instance, that the peer-assessment is narrow banded in comparison to the tutor's assessment.

Note that the same area plot shows some agreement if we look into other aspects. For instance, the statistical mode matches at 8 for both the lecturer and the peers. Other metrics that seem to agree are the average and the "total marks given" (figure 2) (even though both are slightly larger in the peer-assessment case). The "total marks given", being a summation, can be seen as a measure of the willingness to give better marks. The big difference is on the standard deviation, which is practically twice as much for the tutor. This is because the tutor marks are broad banded in comparison to the peer assessment marks.
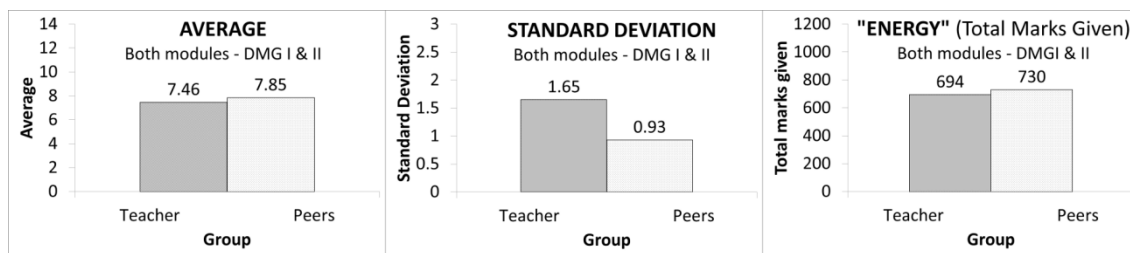
210

211   Figure 2. Generic metrics of the results: average, standard deviation and "energy".

212

213   It might be expected that, under different circumstances – for example, higher

214   stakes assessments with *category-based criteria* defined - the average and mode would

215   change and shift. In that case, the teacher's assessment would probably be smaller

216   (shifted to the left), as observed by other authors, such as Kirby and Downs (2007)

217   when discussing a case-study on self-assessment.

218   *Qualitative analysis between the lecturer marks and the peer-assessment marks*

219   First, it must be noted that we deliberately refrained from using a quantitative rigorous

220   statistical analysis of the results, as found in some other studies, because we found it

221   might conceal some of the particular causes for the peer-assessment outcomes that we

222   are looking for. Van Zundert et al. (2010) pointed out that most of the published works

223   in the literature about peer-assessment, although providing useful insights regarding

224   best practices, are inconclusive with respect to cause-effect relations involved in the

225   peer-assessment process.

226   Results on figure 1 suggest that the data can be divided into 2 main groups. We

227   decided to categorise them into "divergent" and "convergent", as we think these words

228   are the most adequate to define how the perception differs from the tutor to the student.

229   The divergent group, in this study, is proposed to be composed by the grades ranging

230   from 1 to 6 (and possibly 10), with zero correlation between the lecturer and the peers,

11

231 as opposed to the convergent group ranging from 7 to 10 (or 7 to 9) where data seems to

232 present a certain degree of correlation. Divergence occurs in the 1 to 6 marks range as it

233 sounds evident that the 6 students receiving a pass from their peers (5 and 6 marks) are

234 the same 6 receiving a fail from the lecturer (1 to 4 marks).

235   To explain what causes divergence between the tutor and the students, selected

236 parts of the plot represented on figure 1 are discussed below:

237 • Range 1 – 4: This is the range of marks corresponding to a fail. According to the

238   lecturer, 6 students fell into this category. However, the students appear to have

239   avoided failing their peers and did not place any into this category, even though

240   some presentations were quite poor. These results are divergent.

241 • Range 5 – 6: All the 6 students who received a fail from the lecturer (mentioned

242   in the category above) were put in this category by their peers, i.e. not a fail.

243   Nevertheless, the students demonstrate the ability to rank the presentations: the

244   two students that got 1 and 3 from the tutor were given a 5 (pass) by their peers

245   and the four students that got 4 from the tutor were given a 6 by their peers

246   (slightly more than a pass). These results are divergent.

247 • Range 7 – 8: 72% of the marks given by the students in peer-assessment fall into

248   this range while 50% of the marks given by the lecturer fall into this category.

249   These results can be said convergent.

250 • Range 9 – 10: There seems to be a tendency to invert the situation for the

251   highest marks: the students appeared less likely to give a 10 (6 given by the

252   lecturer against only 2 in peer-assessment). Considering the "9" marks, an equal

253   number of "9"'s were awarded by students and tutors. These results can be said

254   convergent, but with a tendency to diverge.

255   Concerning the first two ranges above, in which the results are divergent

256   between the lecturer and the students, peer-assessment was found to be more generous

257   than the tutor's. However, it is should be noted that the students are able to grade the

258   presentations according to their merit - exactly as perceived by the tutors. All concur

259   that there are 6 presentations worse than the others, in which 2 are clearly worse than

260   the other 4, and this was not a subjective judgement: there was unanimous agreement on

261   this.

262   On the other hand, the last range (9 to 10 marks) suggests that students seem to

263   avoid marking their peers much higher than what they feel they are likely to get

264   themselves.

265   Another relevant aspect that is generally pointed out as a source of bias is

266   hostility (Falchikov 2003). However, when dealing with a large number of students and

267   admitting hostility exists only between a few students, this will be averaged out

268   eventually.

269   What is left is a bulk of marks allocated by the students in the tight and narrow

270   7-8 range, even though we must recognise the tutors also felt that the majority of

271   presentations would fit in this category.

272   *Normalisation of the peer-assessment marks*

273   One important conclusion from the previous analysis is that the peer-assessment marks

274   are narrow-banded when compared to the tutor marks. Peer-assessment marks range

275   between 5 and 10 (6 marking intervals considered) whereas tutor marks range between

276   1 and 10 (10 marking intervals considered). In an attempt to stretch this band (i.e., to

277   increase its range) so as to match the tutors' marking range, the following formula is

278   suggested:

279 $$s_{norm,i} = r \cdot s_i - c \qquad (1)$$

280   In this equation, $s_{norm,i}$ is the normalised peer-assessment mark for student i, s_i is

281 the original peer-assessment mark for student i, c is a constant used to limit the

282 maximum marks given and r is the range correction factor given by.

283 $$r = \frac{\Delta T + 1}{\Delta S + 1} \qquad (2)$$

284   with ΔT the tutors' marking range ($\Delta T = T_{max} - T_{min}$) and ΔS the students'

285 marking range ($\Delta S = S_{max} - S_{min}$). In this case, $S_{max} = T_{max} = 10$, $T_{min} = 1$ and $S_{min} = 5$.

286   Determination of c can be done by imposing one chosen criterion, for instance

287 the maximum mark allowed or the target average. In this case, the criterion adopted was

288 that the average would be the same for both the students and the tutor, i.e.:

289 $$c = \frac{1}{N} \sum_{i=1}^{N} (r \cdot s_i - t_i) \qquad (3)$$

290   The values of s_i and t_i are the marks received by student i from their peers and

291 the tutor, respectively. These values are the exact averaged values and not rounded

292 values as shown in the histogram earlier. In the present study, the values so obtained

293 were $r = 1.667$ and $c = 5.564$.

294   The correlation between two sets of marks, say the peer-assessment marks and

295 the tutor's marks, can then be determined from the *Pearson product-moment*

296 *correlation coefficient*:

297 $$\rho_{s,t} = \frac{E\left[(s - \bar{s})(t - \bar{t})\right]}{\sigma_s \sigma_t} \qquad (4)$$

298   where $E[.]$ denotes the ensemble average value of $[.]$ and σ_s and σ_t represent the

299   standard deviation of the peer-assessment marks and tutor's marks respectively.

300
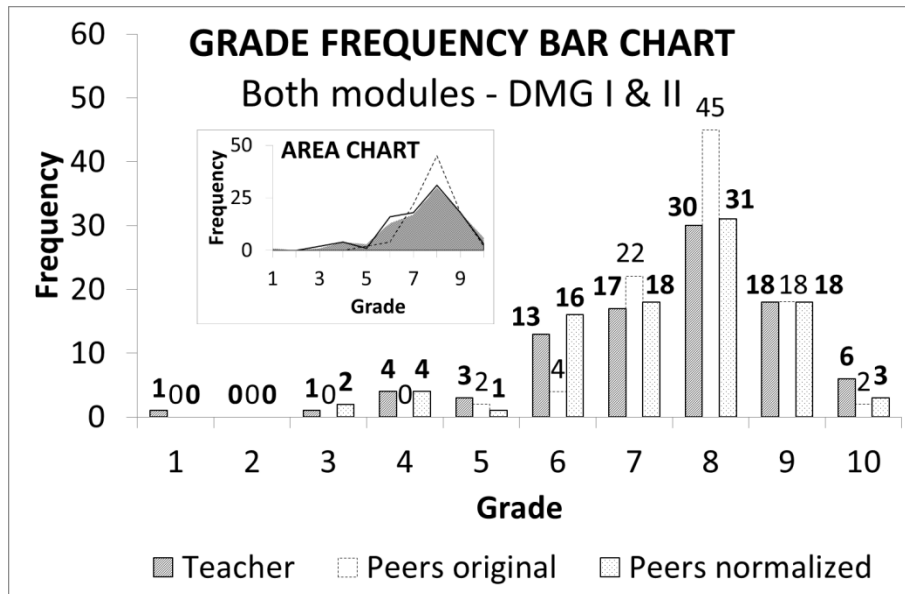
301

302   Figure 3. Histogram showing the amount of times a grade was given, including

303   normalisation of the peer-assessment marks.

304

305   Results of the normalisation procedure are shown in Figure 3 bar chart, along

306   with the tutors' marks and the original peer-assessment marks. It is very interesting to

307   notice how this adjustment process changed data. Now there is a very strong correlation

308   between the two sets of data (tutor and peers-normalised), highlighted by the area charts

309   that are almost perfect matches. In terms of the Pearson correlation factor between the

310   peer and tutor marks (equation 4), it improved from 94% to 99% after normalisation.

311   Analysis of these results will be done following the same reasoning as before, i.e. by

312   highlighting selected ranges of the plot represented on figure 3:

313   • Range 1 − 4: The same 6 students that were identified by both the students and

314   the tutors as having the worst presentations are now in the 'fail' range (peer-

315   assessment originally placed them at the 5-6 range). However, even though the

316   normalised range was stretched so as to include 1, the worst mark obtained was

317   as high as a 3. It is interesting to note that, now, only one student is divergent

318   between the tutors' perception and the students' perception.

319   • Range 5 – 8: All results are now very similar when comparing the tutors and the

320   normalised peer-assessment marks, especially with respect to the 6th and the 8th

321   marking ranges, in which the relative difference between the number of students

322   included by the tutors and the peers drops from 69% to 23% in case of the 6th

323   range and drops from 50% to only 3% in case of the 8th range.

324   • Range 9 – 10: There was some improvement in this range as well. This

325   normalisation allowed that one student being graded 9 by their peers now

326   receives a 10 and one student being graded 8 by their peers now receives a 9.

327   Finally, a comment on the apparent coincidence between the lecturer's marks

328 and the normalised peer-assessment marks. At a first glance, it may seem that equation

329 (1) and the subsequent reasoning serves only to adjust the numbers so as to fit the

330 tutor's marks. Nevertheless, this is only valid happen because the qualitative judgement

331 of the students is coherent with the tutor's judgement. What is happening is that the

332 different presentations were naturally sorted from the worst to the best because the

333 criteria (being *holistic*) tended to be based on comparison.

334   Conclusion

335 In this study, an attempt has been made to better understand how well peer- and tutor

336 assessment correlate, even if the awarded marks differ significantly initially.

337 Engineering undergraduate students were invited to get involved in peer-assessment by

338 marking one another's presentations about their Technical Drawing and Geometrical

16

339   Modelling projects. A set of marking data with statistical significance (93 students'

340   sample) was analysed to assess how differently the students mark their peers when

341   compared to the tutors.

342   The peer-assessment was *blind* – the marks were kept anonymous from the

343   students - to reduce any fears of discrimination. However, it was not possible to

344   implement it in a double-blind peer-review fashion, as the students knew who they were

345   marking, so it is possible that fear for discrimination and hostility were not totally

346   reduced, although hostility is expected to be very localised and, thus, averaged out.

347   As might be expected, students tend not to fail their peers – neither do they tend

348   to award the highest marks either. This means that the band where the students give

349   their marks tends to be narrower than the one used by the lecturer. Yet, it was found that

350   they are able to make a correct relative judgement between their peers' performances

351   (and thus themselves).

352   To mitigate this, a normalisation procedure to adjust peer-assessment marks has

353   been presented, in an attempt to correlate peer- and tutor assessment in quantitative

354   judgements. This adjustment, that proved to significantly improve the correlation

355   between the tutor's and the students' assessments, is still very dependent on the

356   teacher's perception (e.g., marking range and average). Thus, further study is still

357   needed so as to conclude about its practical usefulness, especially when tutor's marks

358   are not available to establish reliable normalisation parameters.

359

360   **References**

361   Boud, D., and Falchikov, N. (2006) Aligning assessment with long term learning.

362   *Assessment and Evaluation in Higher Education*, 31(4), 399–413.

363   Brown, S., Sambell, K., and McDowell, L. (1998) What do students think about peer

364   assessment? In Peer Assessment in Practice, (ed) Brown, S., 107–112.

365   Birmingham, UK: Staff and Educational Development Association (SEDA),

366   paper 102.

367   Chuck, J.-A., and Young, L. (2004) A Cohort-Driven Assessment Task for Scientific

368   Report Writing. *Journal of Science Education and Technology*, 13(3), 367-376.

369   Cotner, S. H., Fall, B. A., and Wick, S. M. (2008) Rapid Feedback Assessment

370   Methods: Can We Improve Engagement and Preparation for Exams in Large-

371   enrollment Courses? *Journal of Science Education and Technology*, 17, 437-

372   443.

373   Dochy, F. J. R. C., and McDowell, L. (1997) Assessment as a tool for learning. *Studies*

374   *in Educational Evaluation*, 23, 279-298.

375   Earl, L. M. (2003) *Assessment as Learning*. California: Corwin Press Inc.

376   Falchikov, N., and Boud, D. (1989) Student self-assessment in higher education: a

377   meta-analysis. *Review of Educational Research*, 59(4), 431–470.

378   Falchikov, V. (2003) Involving students in assessment. *Psychology Learning and*

379   *Teaching*, 3(2), 102–108.

380   Gipps, C. V. (1994) *Beyond Testing: Towards a Theory of Educational Assessment*.

381   London: Falmer Press.

382   Hatzipanagos, S., and Rochon, R. (2010) Introduction to the special issue: approaches to

383   assessment that enhance learning. *Assessment and Evaluation in Higher*

384   *Education*, 35(3), 491–492.

385   Hunter, D. (1999) Developing peer-learning programmes in music. *British Journal of*

386   *Music Education*, 16, 51–63.

387   Kirby, N. F., and Downs, C. T. (2007) Self-assessment and the disadvantaged student:

388   potential for encouraging self-regulated learning? *Assessment and Evaluation in*

389   *Higher Education*, 32(4), 475–494.

390   Norton, L. (2004) Using assessment criteria as learning criteria: a case study in

391   psychology. *Assessment and Evaluation in Higher Education*, 29(6), 687–702.

392   Orsmond, P. (2011) *Self- and Peer-Assessment: Guidance in Practice in Biosciences*.

393   Leeds: UK Centre for Bioscience, The Higher Education Academy.

394   Russel, M., Haritos, G., Combes, A. Individualising students' scores using bling and

395   holistic peer assessment. *Engineering Education*, 1, 50-60.

396   Silva, J. M. M., and Maia, N. M. M. (1997) *Theoretical and Experimental Modal*

397   *Analysis*. Somerset, UK: Research Studies Press Ltd.

398   Tsivitanidou, O. E., Zacharia, Z. C., and Hovardas, T. (2011) Investigating secondary

399   school students' unmediated peer assessment skils. *Learning and Instruction*,

400   21, 506-519.

401   Van Gennip, N. A. E., Segers, M. S. R., and Tillema, H. H. (2010) Peer assessment as a

402   collaborative learning activity: The role of interpersonal variables and

403   conceptions. *Learning and Instruction*, 20, 280-290.

404   Van Zundert, M., Sluijsmans, D., and Van Merriënboer, J. (2010) Efective peer

405   assessment processes: research findings and future directions. *Learning and*

406   *Instruction*, 20, 270-279.