

Exploring the Reliability of Self-Assessment and Peer-Assessment in Oral Presentations in Economics: A Sample of Postgraduate Students at a South African University

Josue Mbonigaba^{1*} and Saidou Baba Oumar^{2**}

¹University of KwaZulu-Natal, Durban, KwaZulu Natal, South Africa

²University of Buea, Cameroon

E-mail: * <mbonigaba@ukzn.ac.za>, ** <saidoubo@gmail.com>

Address for correspondence

Josue Mbonigaba

Box 61133 Bishopsgate Durban

South Africa

Telephone: +27312602503

Fax 0312607871

E-mail: mbonigaba@ukzn.ac.za

ABBREVIATED TITLE MARKING INDIVIDUAL CLASS PRESENTATION AT A UNIVERSITY

KEYWORDS Economics, postgraduate, students, South Africa

ABSTRACT The evidence about the reliability of self-assessment and peer-assessment marks in higher education has not been established. This paper contributes to the literature by exploring the reliability of such assessments in oral presentations in a sample of postgraduate students at the University of KwaZulu-Natal (UKZN), South Africa. Data were collected as part of a teaching and learning exercise for a period of 7 years (2007-2013) from successive cohorts of students. The sample consisted of 238 students. The reliability of these assessments was explored in relation to assessments by the lecturer based on the fact that the

lecturer is more specialised in assessing students' work. The analysis of the data is conducted with descriptive and inferential statistics. In light of this study's evidence, peer assessment marks and self-assessment marks in oral presentations are different from the lecturer's marks in most respects and are therefore not reliable. We therefore conclude that these assessments are not reliable at least for this type of assessment, for this particular sample and for the course assessed. Given the importance of self-assessment and peer-assessment in higher education, more research is recommended, particularly in developing countries, to find conditions under which these assessments could be reliable.

I. INTRODUCTION

Assessment has been found in the literature to influence the way in which students learn (Scouller 1998). In this respect, the use of peer assessment and self-assessment has been increasingly being considered important in higher education literature (Dochy, Segers, and Sluijsmans 1999; Boud, and Falchikov 2007; Spiller 2012), because of their prospects for engaging students in exploiting their full potential in learning (Spiller 2012). Peer and self-assessment not only increase learning potential, but also achieve other purposes, such as developing a cooperative spirit that is useful to the professional careers of students after graduation (Falchikov and Goldfinch 2000; Boud and Falchikov 2006). However, peer and self-assessment have been mainly used in formative assessment and they have been found to do better in this respect (Nicol 2006; Galbraith, Hawkins, Holmboe 2008, Nicol 2009) than using them in the awarding of marks (Kirby and Downs 2007: 490, Spiller 2012:13). Consequently, the evidence of the reliability of the scores arising from these assessments in summative assessment has not been extensive because of difficulties in administering these assessments (Yorke 2007). This paper explores such reliability by comparing the marks awarded by these two types of assessment with the marks given by the lecturer in a postgraduate course at the University of KwaZulu-Natal (UKZN) in South Africa.

Peer assessment, like any other type of assessment, involves determining the expected standards of a course of study and using tools that enable making a judgment about whether or not these standards are achieved. Peer assessment consists of feedback from peers regarding the quality of work produced with respect to expected standards and criteria. Usually peer assessment is most relevant for formative assessment in which, for instance, drafts of group work are circulated to other groups in the class, and each group takes turns in reporting feedback on their peers' work in relation to preset criteria.

Peer assessment is considered very important because it conforms to contemporary approaches to learning in higher education: students engage actively in learning in the context of constructivist learning and teaching practices. Peer assessment has been considered to be advantageous over traditional approaches of assessment in which the power of assessment has been in the hands of the lecturer. Through peer assessing, students develop critical thinking, learn more than what is expected, and become confident and deep learners. Peer feedback based on preset and well defined criteria enhances collaboration in learning and interchanges about what constitutes good work, therefore providing the lecturer and learners alike with insights into areas that need improvement. Peer assessment facilitates future success in the job market (Sadler, 2005; Boud and Falchikov, 2006; O'Donovan, Price, and Rust, 2008) and helps students grasp the complexity of professional judgment (Ball, Bew, Brown, *et al.*2012). Peer assessments have been used mainly in formative assessment and therefore not for the allocation of marks. It is pointed out in the literature that although peer assessment can be used to award marks, the process has been fraught with difficulties in most cases (Spiller 2012)

Like peer assessment, self-assessment has increasingly been considered to be a way to engage students in the process of learning (Lawson, Taylor, Thompson *et al*, 2012). Self-assessment consists of a process through which students make a judgment on their own achievements based on defined standards. In this respect, self-assessment allows students to develop a critical approach (Lew, Alwis, Schmidt 2010). towards their own work with respect to the achievement of expected standards. Andrade and Du (2007: 160) define self-assessment as a process through which students reflect on and evaluate the quality of their work so that they can identify strengths and weaknesses and revise their work accordingly. Furthermore, self-assessment allows the students to develop knowledge beyond what the teacher gives them. The literature indicates that self-assessment can help not only to solve a problem conceived by the lecturer, but also to solve the problems unconceived by the lecturer (Brew 1995: 57). Self-assessment has been beneficial in that students constantly refer to the performance criteria set up and therefore are constantly exposed to opportunities to improve their work. While self-assessment is important in this respect, its successful implementation requires that students understand clearly what is expected of them, know how standards are measured based on clearly defined criteria, and know that cheating and collusion can be detected and penalized (Boud 1995: 182; Sadler 2005; Boud, and Molloy 2012).

Using self-assessment in awarding marks has been done in a variety of ways. The literature indicates that the final mark can be allocated by the teacher after combining the self-assessment mark and the teacher's mark, or by using the teacher's marks and adding a percentage reflecting the quality of self-assessment (Spiller 2012). Another approach to self-assessment consists of giving students a model answer and asking them to indicate the differences between the model answer and their own responses and to allocate the marks. Regarding the reliability of the marks from self-assessment, this has not been firmly established. A recent research study indicated that marks from self-assessment converge with those of the tutors over time for strong students, though not for weak students (Boud, Lawson, Thompson 2013:941).

Despite difficulties in using peer assessment and self-assessment in awarding marks even with preset criteria (Sadler 2009), such assessments can be important in saving time spent on marking by lecturers and freeing up time for research purposes. If they are found reliable, they could be implemented more regularly to take advantage of their benefits as highlighted above. It is therefore important to start looking into how these assessments can be used in marking. The starting point in this process is to assess whether the resulting marks are reliable measures by comparing them to more standard measures such as lecturer's marks. Given that the evidence is not widespread and seems to be context-specific, it is important that the evidence for the reliability of peer assessment and self-assessment continues to be sought. To this end, this paper seeks to explore their reliability in a specific assessment task, this being a class oral presentation. Specifically, it attempts to answer the question of how reliable self-assessment and peer assessment marks are in measuring achievement of preset standards in class oral presentations using a sample of postgraduate students at UKZN in South Africa.

II. METHODOLOGY

To explore the reliability of peer-assessment and self-assessment the study compared, in different perspectives, the marks allocated by these two types of assessment to the marks allocated by the lecturer. Using the marks of the lecturer as a benchmark for reliability was based on the fact that the latter is more specialised in assessing students' work. So the extent of the reliability of the marks from self-assessment and peer-assessment was judged on the basis of their deviation from the lecturer's marks.

In exploring the reliability of self-assessment and peer-assessment, the study used self-assessment, peer-assessment, and lecturer-assessment marks, recorded and kept to this effect. The data were collected as part of a teaching and learning exercise only and marks records from self-assessment and peer-assessment were not recorded as component of the final mark of the course. These data were collected on 238 students, over the period 2007-2013, from seven successive cohorts of postgraduate students enrolled on the course "Economics of Health" at UKZN, South Africa. Table 1 shows the distribution of demographic characteristics of the students' sample.

Insert Table 1 about here

For each cohort, a student was asked to choose a topic among the 12 topics to be covered over the 12 weeks of the semester. Since in every cohort the number of students was greater than the number of topics, some students had to work on the same topic albeit with individual oral presentations. The presenting student was then asked to self-assess their own oral presentation after presentation while the rest of the students were asked to peer-assess the work during its presentation. All three types of assessment were done against preset criteria, specified for each of the following intervals of performance: 'fair (20-39%)', 'good (40-59%)', 'very good (60-79%)', and 'excellent (80-100%)' (See details in the Appendix).

Two types of analysis were used, namely descriptive statistics and inferential statistics. In descriptive statistics, reliability of self-assessment and peer assessment marks was explored by examining, in relation to the lecturer marks, placement of the cohorts in interval of performance, variability of marks, average marks of different cohorts, and trends of marks from one cohort to another. So, even if independently, each presenting student was placed in

an interval of performance and given a mark by each fellow student (peer assessment), by his or herself (self-assessment) and by the lecturer (lecturer assessment), descriptive analysis was conducted at cohort level at the sample level using average values.

The second analysis consisted of formal tests on data for the whole sample (period 2007-2013) to conclude whether the marks from the lecturer were indeed statistically different from peer assessment and self-assessment marks. In this respect, two tests were performed, the t-test and the Wilcoxon signed-rank test. The t-test was used to compare the marks of the lecturer to the marks of the peer assessment, while the Wilcoxon signed-rank test was used to compare the marks of the lecturer and the marks of the self-assessment. Also, the Wilcoxon signed-rank test was used to compare the marks from peer-assessment to the marks from self-assessment.

The use of the t-test in comparing two populations requires the populations to be normally distributed and have equal variances. These conditions were fulfilled for peer assessment marks and lecturer marks, and therefore the t-test were used in their comparison. In contrast, the normal distribution requirement of the t-test was not satisfied for the self-assessment marks, which were skewed to the left. For skewed data, the literature suggests the use of the Wilcoxon signed-rank test and this test was used in all comparisons of marks that involved self-assessment marks.

For the whole sample, the study tested to see whether the lecturer marks were on average statistically lower than peer assessment or self-assessment marks. This one-tailed test was based on the information from descriptive statistics that the lecturer's marks were lower than the marks from self-assessment and peer-assessment. Two tests were conducted. The first (t-test) consisted of testing whether the lecturer's marks were statistically lower than peer assessment marks, and the second (Wilcoxon signed-rank test) consisted of testing whether the lecturer's mark were statistically lower than self-assessment marks. The paper conducted a third test (Wilcoxon signed-rank test) to test whether the peer assessment marks were statistically lower than self-assessment marks.

The reliability of peer and self-assessment marks was also analysed by testing how the marks compared to the lecturer's marks in sub-groups of students. This analysis is motivated by the suspicion that peer assessment might bias groups' marks depending on how the groups of assessed students relate with the groups of student peer assessors. For instance, we suspected that groups of student from a given social group might be favoured by groups of students from the same social group. Furthermore, cheating and collusion have been found to be

possible in these types of assessment (Boud, 1995: 182; Sadler, 2005; Boud, and Molloy, 2012). Therefore, the paper tested to see whether there was bias in the marks assigned to specific groups of students, who were stratified by gender and race groups to understand whether their marks were higher or lower than the lecturer's marks. The reliability of the marks was questioned if the results in these groups behaved differently from the overall results. The null hypothesis in each testing case was: H_0 : the average lecturer's mark is equal to the average marks of a given type of assessment against the alternative: H_a : the average lecturer's mark is lower than the marks of a given type of assessment. All analyses were conducted using STATA and Excel software. The results of these analyses are reported next.

III. RESULTS

In exploring the reliability of self-assessment and peer-assessment, we start by examining, in relation to the lecturer's marks, how these two types of assessment place the cohorts' average marks in interval of performance. Table 2 presents such results.

In Table 2, "x", shown against each type of assessment and under a given interval of performance, means that the marks from that type of assessment place the cohort in that interval of performance. Based on this understanding, one can see that self-assessment marks deviate from the lecturer's mark. Table 2 shows that the self-assessment marks and the lecturer marks have no cohort of students in common in any interval of performance. In contrast, Table 2 shows that peer-assessment marks and lecturer's marks place some cohorts (2007, 2010, and 2013 cohorts) in the same interval of performance. Furthermore, the results suggest that self- assessment marks place cohorts in higher intervals of performance than do lecturer's marks. Note however that when the whole sample is considered, lecturer's marks and peer assessment place the sample in the same interval of performance. Because self-assessment and peer assessment marks do not place all cohorts in the same interval of performance suggested by the lecturer's marks, it can be said that their marks differ more generally to the lecturer's marks.

Insert Table 2 about here

The results in Table 2 show also the variability of marks around the average marks for each type of assessment. Variability of marks, measured by standard deviation, reflects how differently members of the cohort perform. Given that this variability across the three types of assessment is explored on the same cohort or sample, variability of marks reflects variability of these types of assessment in allocating the marks. Keeping this caveat in mind, the results show greater variability in self-assessment marks than in the marks of the other two types of assessment. The evidence of greater variability in self-assessment and peer-assessment marks in relation to the variability of the lecturer's marks suggest the unreliability of the marks from former two assessments.

Exploring differences in the mean marks on each cohort, results in Table 2 highlights that the average marks of self-assessment and peer-assessment is greater than the average marks of the lecturer assessment. These results apply also to the whole sample although it is important to note that peer-assessment marks are closer to the lecturer's marks and exhibit a more similar trend with lecturer's marks than self-assessment marks do (see Fig.1).

Insert Figure 1 about here

Could this mean that peer assessment marks are reliable and self-assessment marks unreliable? Until formal tests are done, this question cannot be answered at this stage. The reliability of peer- assessment marks and self-assessment marks can only be established based on the evidence as to whether or not the average mark of the lecturer is indeed statistically lower than the average mark from the other two types of assessment. This evidence is presented in Table 3.

Insert Table 3 about here

This evidence from the sample data shows that the average marks from the lecturer are less than the average marks of peer assessment as indicated by a negative t-test statistic ($t = -3.228$). The p-value of 0.0028 suggests that this difference between the lecturer's marks and peer- assessment marks is statistically significant at a 5% level of tolerance, implying that indeed the lecturer's marks are lower than the peer-assessment marks. The same conclusion applies to the other two tests. Specifically, the lecturer's marks are statistically lower than the self-assessment marks on the basis of the Wilcoxon signed-rank test statistic ($z = -3.705$) with a corresponding p-value of 0.0002, and peer-assessment marks are lower than self-assessment marks on the basis of the Wilcoxon signed-rank test statistic ($z = -3.189$) with a corresponding p-value of 0.0014.

To test for the reliability of peer assessment, the study conducted a test to see whether or not the finding for the whole sample was applicable to gender categories of students or whether the trend in the marks of peer assessment or self-assessment followed the trends in the marks of the lecturer in each gender category. Table 4 shows the results.

Insert Table 4 about here

Results in Table 4 show that the lecturer's average mark of 73.73 for female students is lower than the female peer assessor's mark of 78.1 for the same gender, and this difference is statistically significant at a 5% level of tolerance (p-value = 0.0001). Furthermore, the lecturer's average marks of 73.73 for female students is lower than the average marks of the peer male assessors of 76.43 for the same category and this difference is statistically significant at a 10% significant level (p-value 0.078). Since these marks are consistent with the overall marks, there is no gender bias in the allocation of marks. This fact is also

confirmed by the evidence that trends in marks of female peer assessors and male peer assessors behave similarly across gender. So for instance, it is worth noting that female peer assessors allocate more marks to females (78.1%) than they do to males (74.5%). Likewise, male peer assessors allocate more marks to females (76.43%) than they do to males (75.34%). On the basis of these results, it can be concluded that there is no bias in allocation of marks across gender categories, although the results confirm that peer assessments are not reliable since the lecturer's marks are statistically lower than the peer-assessment marks.

In a sample with multi-social groupings, one would expect social group bias in the allocation of marks. To this end, tests were conducted to compare the average marks allocated by the lecturer to each social grouping and the marks allocated to those groupings from peer assessors from different social groups and the trends in marks of a given grouping. The results of such analysis are in Table 5.

Insert Table 5 about here

The lecturer's marks for different race-based groups were generally lower than the marks allocated to the same groups by peer Asian assessors, except for assessing the White students group. This result reveals that Asian students tended to overestimate the ability of race-based groups, except for the White student group. The results of the marks allocated to different race-based groups by the lecturer are generally lower than the marks allocated to these groups by Black peer assessors, except for the Coloured group, suggesting that Black students tended to overestimate the ability of race-based groups except the Coloured student group. The marks allocated by the lecturer to the various race-based groups were higher than the marks allocated to the same groups by Coloured students. This may indicate that the Coloured group tended to underestimate the ability of race-based groups except for themselves. The marks allocated by the lecturer to these race-based groups were higher than the marks allocated to these groups by the White students, and this may imply that White students underestimated the intellectual ability of other race-based groups.

Analysing the trends in the marks as one moves across these racial groupings, the results show that Asian assessors rank the groups, from lower to higher, in this order, Black, White, Coloured and Asian, while the lecturers rank them in the order, Coloured, Black, White and Asian. Black assessors rank groups in this order, Coloured, White, Black, and Asian, while the lecturer ranking is in the order of White, Black, Coloured, and Asian. Since the ranking

of race-based groups in terms of marks is not consistent with the lecturer's ranking, these results indicate that there might be racial bias in the marks assigned by peer assessment.

Briefly, the evidence presented in the results for race-based groups points not only to inconsistencies between the lecturer's marks for the different groupings and the marks allocated by peer assessment to these groupings, but that they are also inconsistent with the overall marks trends of the sample. For instance, not all the average lecturer's marks for the different groups were statistically lower than the marks allocated by peer assessors, highlighting that some peer assessors overestimated the ability of students from different groups. This result deviates from the results observed for the whole sample. Although the nature of the bias could not be detected, the results highlight that peer assessment may carry race-based group bias in this sample of students

IV. DISCUSSION

The benefits of self-assessment and peer-assessment in higher education are twofold: freeing up marking time for research purposes and taking pedagogical advantages of these two types of assessment. As one of the teaching activities that consumes much time is marking students' work, reducing marking time by using other methods, such as peer assessment and self-assessment could be one of the avenues to increasing time dedicated to research and consequently one way to increase research output. Furthermore frequent uses of self-assessment and peer-assessment in summative markings could allow students to adopt learning styles conducive to deep learning and relevant to their future careers activities. Before peer assessment and self-assessment are substituted for traditional lecturer marking however, the marks produced by these types of assessment need to be ascertained as being reliable measures of the intellectual ability of students.

In a bid to contribute to the establishment of this evidence, this study explored the reliability of peer assessment and self-assessment by exploring their reliability in oral presentations in a postgraduate course at UKZN South Africa. The reliability of self-assessment and peer-assessment was judged with respect to the extent, to which they were similar to the lecturer's marks, the latter being used as benchmark.

The main findings of the investigation were that the marks allocated from peer and self-assessment in this sample were generally higher than the marks allocated by the lecturer, both at the cohort level and sample level. Both peer and self-assessment marks overestimated the marks in oral presentations, in relation to the marks estimated by the lecturer. Peer assessment marks, although also overestimating students' marks in oral presentations trended similarly to the lecturer's marks than did self-assessment (Fig.1). The reliability of peer assessment was further explored by focusing on whether or not there was a bias in the allocation of marks by assessors in a specific social group, such as gender groups, as they assessed males and females in the sample. The main finding in this respect was, that although marks are overestimated by groups of males and females, there was no bias in the marks males allocated to females and vice versa, on the basis that trends in these mark were consistent with the trends in marks allocated by the lecturer (Table 4).

The reliability of peer assessment was further investigated in order to establish whether there is bias in the allocation of marks by social groups of assessors based on race, namely, Asian, Black, Coloured and White, for these four race-based groups (Table 5). The marks from the race-based group peer assessors of other race-based groups were compared with the lecturer's marks for these four groups. While the trend in the marks allocated to male and females assessed by groups of male and female assessors were similar to trends in the lecturer's marks, a mixed picture (overestimation or underestimation) of peer assessment across race-based groups was found, showing additional bias in mark allocations to different race groups although the nature of the bias in terms of whether there is a specific group favoured or not favoured by other groups was not established.

The findings have important implications with respect to using peer assessment and self-assessment in allocating marks to student work. Specifically, in light of the findings, using these types of assessment in allocating marks to students mean promoting inaccuracy in measuring student oral presentation's marks, which may impact negatively on educational outcomes. Therefore these types of assessment should not be used in allocating the marks to students in postgraduate oral presentations.

The findings in this paper have some precedents. Other studies investigating the accuracy of peer and self-assessment in allocating marks to students have underlined that only under certain conditions; can these forms of assessment result in reliable marks. These studies have argued, however, that such marks still need to be moderated before being recoded as final marks. In this respect this study reinforced what is known by using empirical evidence

specific to oral presentations. By providing the evidence in the developing world, this paper contributed also to the literature in this world. Most of related evidence has only been produced in developed countries.

The evidence presented in this paper implies that the accuracy of peer assessment may be context-specific. In particular, the fact that results of peer assessment were more biased across race-based groups may reflect the past history of the country in which there are perceptions of intellectual inequality across different race groups. These and similar cases could prevail differently in other contexts.

V. CONCLUSION AND RECOMMENDATIONS

At this stage, peer and self-assessment should not be used to substitute traditional lecturer's marking in postgraduate oral presentations until solid evidence of the reliability of these forms of assessment in assigning marks to students' work is established. Even though this study was only on a single piece of assessment, its results were consistent with previous work. Consequently, before these assessments are used in assigning marks that contribute to the final mark of the course, more diverse research is needed. Therefore, a key recommendation is that more studies are conducted in various disciplines, subjects, assessment tasks; using different approach and particularly in developing countries. Furthermore, more research is recommended to find conditions under which these assessments are reliable.

5. ACKNOWLEDGEMENTS

The authors would like to thank in advance the anonymous reviewers for their valuable comments and Carol Brammage for English editing.

7. REFERENCES

Andrade H, Du Y 2007. Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education* 32 (2): 159-181.

Ball S, Bew C, Brown S, Kleiman P, May, H, McDowell L, Morris E *et al.* 2012. *A Marked Improvement: Transforming Assessment in Higher Education*. The Higher Education Academy.

Boud D 1995. *Enhancing Learning through Self-assessment*. London: Kogan Page.

Boud D, Falchikov N 2006. Aligning assessment with long-term learning.” *Assessment & Evaluation in Higher Education* 31(4): 399-413.

Boud D, Falchikov N 2007. *Rethinking Assessment in Higher Education*. London: Kogan Page.

Boud D., and E. Molloy E 2012. Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education*, doi: 10.1080/02602938.2012.691462.

Boud D, Lawson R, Thompson DG 2013. Does student engagement in self-assessment calibrate their judgment over time? *Assessment & Evaluation in Higher Education* 38(8): 941-956

Brew A 1995. What is the scope of self-assessment? In D. Boud (Ed.). *Enhancing Learning through Self-assessment*, 48-63, London: Kogan Page. pp 48-63

Dochy F, Segers M, Sluijsmans, DMA 1999. The use of self-, peer and co-assessment in higher education: A Review.” *Studies in Educational Evaluation* 24(3): 331–350.

Falchikov N, Goldfinch J 2000. Student peer assessment in higher education: A Meta-analysis comparing peer and teacher marks. *Review of Educational Research* 70(3): 287-322.

Galbraith RM, Hawkins RE, Holmboe ES 2008. Making self-assessment more effective. *Journal of Continuing Education in the Health Professions* 28 (1): 20–24.

Kirby NF, Downs CT 2007. Self-assessment and the disadvantaged student; potential for encouraging self-regulated Learning?” *Assessment & Evaluation in Higher Education*, 32(4): 475-494.

Lawson RT, Taylor DG, Thompson L, Simpson M, Freeman L, Treleaven M, Rohde F 2012. Engaging with graduate attributes through encouraging accurate student self-assessment. *Asian Social Science* 8 (4): 3–12.

Lew MDN, Alwis WAM, Schmidt HG 2010. Accuracy of students’ self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education* 35 (2): 135–156.

Nicol DJ, Macfarlane D 2006. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education* 31 (2): 199–218.

Nicol DJ 2009. Assessment for learner self-regulation: Enhancing achievement in the first year using learning technologies. *Assessment & Evaluation in Higher Education* 34 (3): 335–352.

O'Donovan B, Price M, and Rust C 2008. Developing student understanding of assessment standards: A nested hierarchy of approaches. *Teaching in Higher Education* 13 (2): 205–217.

Scouller K 1998. The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education* 35: 453–472.

Sadler DR. 2005. Interpretations of criteria-based assessment and grading in higher Education. *Assessment & Evaluation in Higher Education* 30: 175–194.

Sadler DR 2009. Indeterminacy in the use of preset criteria for assessment and grading in higher education. *Assessment & Evaluation in Higher Education* 34: 159–179.

Spiller D 2012. *Assessment Matters: Self-Assessment and Peer Assessment*. Teaching Development| Wāhanga Whakapakari Ako. Available at http://www.waikato.ac.nz/tdu/pdf/booklets/9_SelfPeerAssessment.pdf, (Retrieved December 10, 2013)

Yorke M 2007. *Grading student achievement in higher education: Signals and shortcomings*. London: Routledge.

8. TABLES

Table 1: Summary of students' sample characteristics

Social group	Male	Female	Total
Black	13	42	55 (23%)
Coloured	4	17	21 (9%)
Asian	85	56	141 (59%)
White	10	11	21 (9%)
Total	112 (47%)	126 (53%)	238 100%

Table 2: Comparison of self-assessment and peer-assessment marks to the lecturer's marks

Cohorts	Type of assessment	Average mark	Standard deviation	Level of performance			
				80-100	60-79	40-59	20-39
2007	Per cohort						
	Self-assessment	81	16	x	-	-	-
	Peer-assessment	75	9	-	x	-	-
2008	Lecturer assessment	68	6	-	x	-	-
		85	17	x	-	-	-
	Self-assessment	73	10	-	x	-	-
2009	Peer-assessment	64	8	-	-	x	-
	Lecturer assessment	83	16	x	-	-	-
		76	10	-	x	-	-
2010	Self-assessment	70	7	-	-	x	-
	Peer-assessment						
	Lecturer assessment	86	12	x	-	-	-
		70	9	-	x	-	-

2011	Self-assessment	62	6	-	x	-	-
	Pee-assessment	83	15	x	-	-	-
	Lecturer	77	9	x	-	-	-
	assessment	71	6	-	x	-	-
2012	Self-assessment	85	16	-	x	-	-
	Pee-assessment	74	10	-	x	-	-
	Lecturer	65	7	-	-	X	-
	assessment						
2013		84	16	x	-	-	-
	Self-assessment	75	12	-	x	-	-
	Pee-assessment	67	5	-	x	-	-
	Lecturer						
Sample	assessment						
	Self-assessment	84	17	x	-	-	-
	Pee-assessment	74	10	-	x	-	-
	Lecturer	67	8	-	x	-	-
	assessment						
	For the whole						
	sample						
	Self-assessment						
	Pee-assessment						
	Lecturer						
	assessment						

The sign “-“ means no type of assessment marks place a cohort of students in that interval of performance

Table 3: Comparison of peer assessment marks, self-assessment marks with the lecturer assessment marks

Lecturer- peer assessment			Lecturer- self-assessment			Peer assess- self-assessment		
Left tail test			Left tail sign test			Left tail sign test		
Hypotheses	t-stat	p-value	Hypotheses	z-stat	p-value	Hypotheses	z-stat	p-value
Ho: LM-PAM=0 Ha: LM-PAM<0	- 3.22 8	0.00 28	Ho: LM-SAM=0 Ha: LM-SAM<0	- 3.70 5	0.000 2	Ho: PAM-SAM=0 Ha: PAM-SAM<0	- 3.1 89	0.00 14

LM: The lecturer's mark, PAM: peer assessment mark, SAM: self –assessment mark, N/A: the test is not applicable in this case. Sources: Estimates obtained with STATA analyses.

Table 4: Comparison of three types of assessment per gender

Assessed gender	Average mark					
	Female peer assessors		lecturer's mark		comparison	
	lecturer's mark		comparison		Male peer assessors marks	
	marks	z	z	p-value	marks	z
Female	78.1		73.73	-3.798,	76.43	73.73 -3.125
Male	0.0001			0.078		
	74.5		68.06	-3.725	75.34	69.80 -3.253
	0.0002			0.000		

z: test statistics for the Wilcoxon signed-rank test. Source: estimates from stats

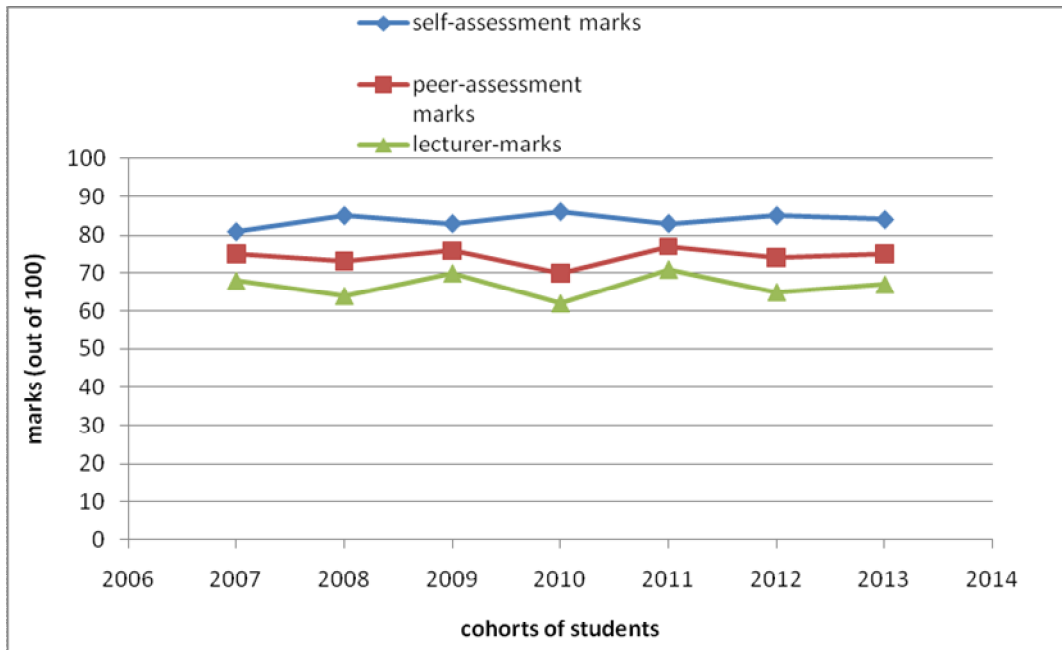
Table 5: Comparison of peer assessment marks of social groups and the lecturer's marks.

Assessed	Average mark							
	Asian	Lecturer	Black	Lecturer	Coloured	Lecturer	White	Lecturer
	p-value		p-value		assessors		p-value	
	assessors		assessors		p-value		assessors	
Asian	83.63	76.02	79.16	74.52	47.40	74.68	72.29	76.67
Black	0.0000		0.000		0.1431		0.4213	
Coloured	71.79	70.89	73.87	69.34	69.9	70.14	69.76	69.76
	0.2430		0.0032		0.7137		1.0000	
White	77.3	60	66.23	69.76	83	64.28	72.25	65
	0.0048		0.6254		0.0178		0.0947	
	73.42	73.71	72.91	66.52	71.87	74.40	70.50	72.66
	0.8190		0.0146		0.5271		0.5282	

Coloured: Born from a marriage between a white person and a non-white person (or their descendants). Elsewhere the term “coloured” as used in South Africa, could be defined as “mixed race”. The study did not distinguish between the nationalities of students. Source: Authors based on STATA estimation.

9.FIGURES

Figure 1: Trends in marks across three types of assessment.



10.APPENDIX

Criteria for assessing presentation

Excellent (80-100%	Very good 60-79%	Good 40-59%	Fair 20-39%
-Discuss a theory	-Discuss theory relevant	-Discuss theory	-Discuss theory

relevant and directly related to the topic -Discuss relevant application of the theory with a focus on few studies (3-4) discussed in depth -excellent coherence and logic in the discussion (intro-to conclusion) -many own critical reflections and opinion about the validity of the theory, relation between results of studies (consistent among themselves and with the theory) -Confident and enthusiastic attitude about your topic and clarity of language	and directly related to the topic -Discuss relevant application of the theory with a focus on few – studies(2-3) discussed in depth -Some coherence and logic in the discussion (intro-to conclusion),presentation clear -some critical reflection -limited confidence, enthusiasm and clarity	relevant and directly related to the topic -Discuss relevant application of the theory with a focus on few - studies discussed in (1-2) depth -some coherence and logic in the discussion (intro-to conclusion, presentation not very clear -very limited or no critical reflection -very limited or no confidence, enthusiasm and clarity	irrelevantly or in a vague way in relation to the topic in question - limited discussion with respect to the application of the theory -inadequate or no coherence or logic in the discussion (intro-to conclusion), presentation unclear -No critical reflection -very limited or no confidence, enthusiasm and clarity
--	---	---	---

List of reviewers

Dr. R. C. Pandey

Associate Professor

Center for Distance Education, North Eastern Hill University (NEHU), Bijnai Complex,
Laitumkhrah, Shillong 793 003, Meghalaya, India

E-mail: rcpandey2912@gmail.com

Dr. Regis Chireshe

Professor

Field of Specialization: Educational Psychology/Guidance and Counseling/Special Needs
Education(Inclusive Education)

Department of Further Teacher Education, P.O.Box 392, UNISA, 0003, South Africa

E-mail: chirer@unisa.ac.za & chireshe@yahoo.co.uk

Dr. Rippen .G.Jassal

Associate Professor

Field of specialization: Human Development

Department of Human Development, Punjab Agriculture University Ludhiana, Punjab, India

E-mail: rippenj@pau.edu, rippenj@hotmail.com

