



The Internet Journal of Allied Health Sciences and Practice

<http://ijahsp.nova.edu>

A Peer Reviewed Publication of the College of Allied Health & Nursing at Nova Southeastern University

Dedicated to allied health professional practice and education

<http://ijahsp.nova.edu> Vol. 7 No. 4 ISSN 1540-580X

Examination of the Inter-Rater Agreement among Faculty Marking a Research Proposal on an Undergraduate Health Course

Shari Rone-Adams, DBA¹

Sandra Naylor, EdD²

1. Lecturer, School of Health Sciences and Social Care
2. Course Director, Senior Lecturer, School of Health Sciences and Social Care

Brunel University
United Kingdom

CITATION: Rone-Adams, S., Naylor, S. Examination of the Inter-Rater Agreement among Faculty Marking a Research Proposal on an Undergraduate Health Course. *The Internet Journal of Allied Health Sciences and Practice*. Oct 2009, Volume 7 Number 4.

ABSTRACT

Purpose: This research looked at inter-rater agreement among faculty marking a research proposal on an undergraduate health course. The purpose of this study was to investigate if there were significant differences between marks as well as the comments given on research proposals that were double marked by two faculty members, where faculty were blind to the other marker's marks and comments. Subjects: The subjects for this study consisted of 80 final year health students who submitted a research proposal and 13 full-time and 10 part-time faculty members who double marked the proposals. Methods: Faculty pairs marked the proposals independently and were required to meet with each other and discuss the marks given, reasons for the mark and the comments. Markers were asked to come to an agreement about the mark and the comments. Analyses: Agreement in marks was assessed using a Bland-Altman plot. Weighted Cohen's Kappa was used to estimate the agreement between the classifications given by markers. Comments were analyzed for differences using thematic analysis. Results: There was a wide discrepancy in the classification of students between markers. The weighted proportions found to agree on classification was 46%. Analysis of the comments indicated a wide discrepancy between markers. Conclusion: The outcomes of this study are similar to previous studies that have looked at inter-rater agreement when double marking was used to mark various types of written assignments. Further exploration of the inter-rater agreement in the marking process and other marking processes that results in a transparent system is needed.

INTRODUCTION

Double marking (of a given assignment) is defined in the literature as a method of marking assignments where scripts are marked independently by two internal faculty who then meet and arrive at an agreed mark.¹ Double marking is an aspect of examining and assessment, by which academic staff try to ensure that transparent and fair mechanisms for marking and moderation are in place that are academically justifiable.² Double marking is a means by which academic staff attempts to produce defensible results. It is essential that assessment judgements are defensible in a growing litigious society where students are increasingly likely to challenge their marks.

Marking of written assignments can be extremely challenging. These challenges arise from the format of written assignments that often preclude close reliance on an evaluation rubric.³ Academics talk about markers needing to have a shared way of thinking in order to systematically agree and disagree on the marks.^{1,2} It has been suggested that double marking will only work in those situations where an open exchange of ideas, values and standards can be shared.² The introduction of reliance on the

markers' knowledge levels or professional judgement as to how the response fits into marking schemes can introduce discretion into the marking procedure and potentially decrease the reliability.

Inter-rater agreement in the double marking of written assignments has not been well researched and much of the research was done some time ago. The research that has investigated the area of double marking of written assignments has often concluded that there was poor inter-rater agreement between markers and that grades appear to often be influenced by chance and specific characteristics of the markers.^{4,5} Several studies in medical schools have concluded that variations among markers in medical courses can be explained by differences in knowledge of students' work.⁶⁻⁸ Another study showed that there was poor agreement between markers marking written case study assignments on a medical undergraduate course.⁹ Markers were general practitioners who taught on the course and university-based staff. The weighted proportions found to agree was 55% and the weighted kappa statistics equalled 0.12 indicating poor agreement between general practitioners and university staff. Poor agreement has been also been shown between markers on oral examinations and portfolios.^{6,7}

Other researchers have shown that there are ways to obtain better agreement between markers. One study looking at written assignments in a physical therapy program showed that with standardized marking schemes reliability between markers increased significantly.¹⁰ It has also been shown that while reasonable agreement between markers was achieved, the reliability of the student marks was only moderate on an undergraduate case study assignment.¹¹

Other research in this area in higher education has also concluded that the resultant final mark is often an average of the marks given by the two markers, even after lengthy discussion about why each marker awarded the marks.^{9,12-14} This calls into question whether the double marking process results in a fair final mark for the student.¹ Theoretically, it could be expected that the discussion between the two markers would result in a systematic analysis of the script identifying where the differences lie between the markers and reconciliation of these differences. There is currently no research verifying that this process works in this manner.

Further research is needed on double marking in other courses. Some research has been done in medical schools, research on other health courses is needed to further explore this area. This research looked at double marking of a research proposal on an undergraduate health course. The purpose of this study was to investigate if there were significant differences between marks as well as the written comments given on research proposals on an undergraduate health course where faculty were blind to the other marker's marks and comments.

METHODS

Research Design

This research looked at the inter-rater agreement between faculty marking a research proposal on an undergraduate health course. Ethical approval was obtained from Brunel University for this study.

Research Proposal Project

In the final year of a three year full time undergraduate health course at a greater London University, students submitted a research proposal during the final term of the program. The proposal was used to assess comprehension of concepts and ideas presented in a research methods course and a literature review course, taken in the second and third years of the program respectively.

Students were introduced to the research proposal in January of the third year of the programme and given two compulsory lectures that explained the requirements and guidelines for submitting the proposal. All materials were posted on the e-learning web site for student and faculty reference throughout the term. Students choose their own topic for the research proposal with guidelines that it must relate to health. Each student was assigned a faculty mentor who guided the student through the development of the proposal. It was not a requirement that the faculty mentor be an expert in the chosen research area as the objective of the proposal was to develop general research skills and not specific topic knowledge. Advice given by faculty mentors normally focused on the following areas: development of the research question, justification of research methods, formulation of a pilot study, referencing, spelling/grammar/syntax. Faculty were allowed to review a one page outline of the proposal. It was the student's responsibility to keep in touch with the faculty mentor as necessary. The pass mark for the proposal was 40%. The marks were grouped into five classifications as shown in Appendix A.

For this study students were asked to submit two copies of the completed research proposal. All proposals were subjected to blind double marking. Blind double marking for this research was defined as two markers, each with a copy of the script who

mark independent of the other marker. Once scripts were individually marked, the markers met and reached an agreed mark and comments for each script.

Markers

Markers for the research proposal were 13 full-time faculty members and 10 part-time faculty members. The number of proposals marked by each faculty member ranged from 4 to 9 and was allocated based on number of hours worked. The markers were also the faculty that acted as research faculty mentors for the students. Each research proposal was randomly allocated (via random assignment to markers as they were submitted) to a particular pair of markers. Markers were paired up randomly and pairs were not consistent between students. No faculty member marked a proposal of a student for whom they had been assigned as the faculty mentor. Support staff held the main list of matching student numbers and names and was responsible for the random allocation of the proposals to the faculty.

Marking Process

The research proposal unit began in January and ended June of the third year of the programme. Eighty research proposals were submitted. Students were asked to submit two copies of their proposal, with only an identifying number, therefore faculty were blind to student names and each faculty member had their own script to mark.

All faculty were formally trained in the use of the classification scheme and marking bands at a staff in-service session prior to the marking of the proposals. Appendix A shows the classification and band definitions and Appendix B shows the data collection form used to document the mark and comments. The philosophy of the team at the time of the marking of the proposals was the belief in the holistic nature of the research proposal rather than the use of a criteria approach. Staff felt that assignment of points to specific parts of the proposal would diminish the judgement of the markers to look at the proposal in a holistic way.¹⁵ Therefore, categories were given for which specific comments should be made but specific points per category were not designated.

All faculty also had access to all information given to the students about the preparation of the proposal throughout the duration of the unit. This information was available to students and faculty on the university's e-learning site.

Faculty had two weeks to complete the marking process. Once faculty had marked the proposals independently, they were required to meet with each other and discuss the mark given, reasons for the mark and the comments. Markers were asked to come to an agreement about the mark and the comments. The form in Appendix B was used by the markers to record their individual comments and marks. These forms were submitted to the researcher for use in data analysis. A final mark and final feedback sheet was also then written that summarized the agreed mark and comments of the two faculty. This sheet was also submitted to the researcher for data analysis.

Data Analysis

To determine inter-marker agreement on classification a weighted Cohen's kappa was used to estimate the agreement between the rating (class) given by the first marker and the second marker. The kappa statistic takes the value 0 when there is no more agreement than would be expected by chance, and is 1 when there is perfect agreement. Agreement is generally considered to be good if kappa is greater than 0.60.¹⁶ This method of analysis has been described for this purpose in previous research.¹¹

Agreement in marks was assessed using a Bland-Altman plot, in which the mean of the first and second markers was plotted against the differences between the two scores.¹⁷ Limits of agreement were computed by calculating two standard deviations above and below the mean difference. These analyses were carried out using Stata version 9.0.¹⁸ This method of analysis has been described for this purpose in previous research.¹¹

RESULTS

The number of students entered into the data analysis was 72. This accounted for eight students where the data collection sheets were not received from the markers. The average age of the students was 20.7 years of age (SD 4.6) upon entrance into the program, with a range of 18 to 42 years of age. Thirty two percent of the students were male with ethnicity consisting of 88 % white, 8% Indian, 2% Chinese, and 2% mixed race.

The markers were 23 faculty members, 13 fulltime staff and 10 part-time staff. The years of experience of the faculty members ranged from 6 months to 20 years, with an average of 7 years of experience in academia. Three faculty members were new to marking research proposals. Pairing of markers for the double marking resulted in 57 different pairs of markers.

Table 1 shows that there is a wide discrepancy in the classification of students between markers. An example of how to read table 1 is as follows:

Looking at Marker 2 under the second category of Very Good, on 14 occasions both marker 1 and 2 agreed on the documents being classified in the Very Good category, however marker 2 also classified one other document in the Very Good category which marker 1 did not and visa-versa, marker 1 classified three other documents in the Very Good category but marker 2 did not.

Table 1: Tabulation of Marker 1 Against Marker 2 in Classification Categories

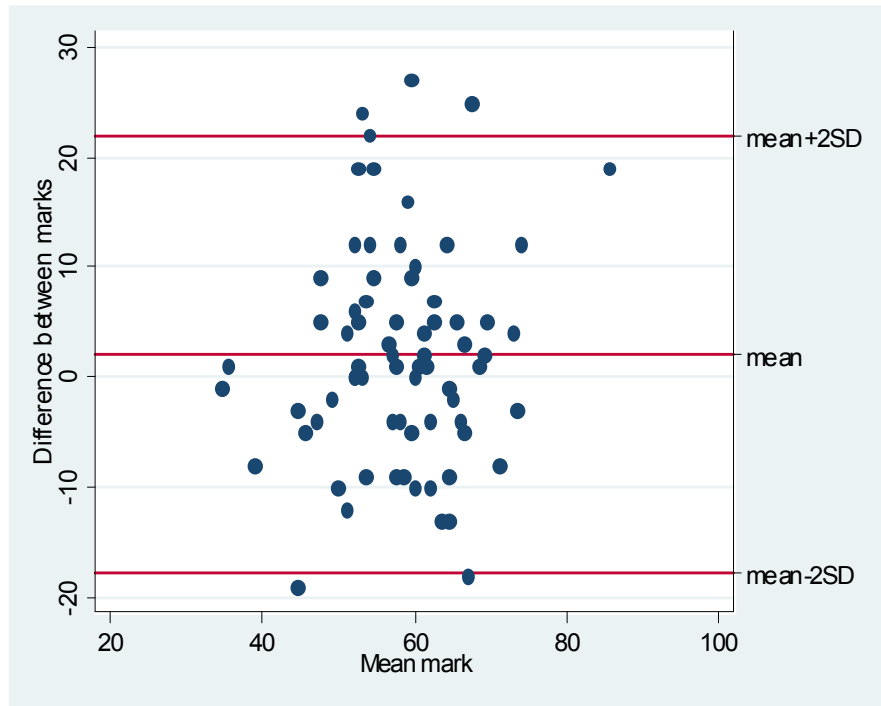
| Marker 1 | Marker 2 | | | | | Total |
|------------------------|------------------------|-----------------------|------------------|------------------------|-----------------|-----------|
| | Excellent (70-100%) | Very Good (60-69%) | Good (50-59%) | Acceptable (40-49%) | Fail (0-39%) | |
| Excellent (70-100%) | 3 | 1 | 3 | 0 | 0 | 7 |
| Very Good (60-69%) | 3 | 14 | 6 | 0 | 0 | 23 |
| Good (50-59%) | 2 | 7 | 11 | 4 | 1 | 25 |
| Acceptable (40-49%) | 1 | 5 | 5 | 3 | 1 | 15 |
| Fail (0-39%) | 0 | 0 | 0 | 0 | 2 | 2 |
| Total | 9 | 27 | 25 | 7 | 4 | 72 |

Note: Bold in the body of the table indicates where markers agree on classification

The weighted proportions found to agree on classification was 46% (33/72). This is shown further in the kappa statistic of 0.34, (0=no agreement, 1=perfect agreement).

In figure 1 x-axis plots the mean marks and the y-axis plots the differences between the marks of the two markers. Figure 1 shows that the mean difference between markers was 2 marks (marker 2 – marker 1), with neither the first nor second marker marking consistently higher or lower than the other. The limits of agreement were -17.86, 21.86 indicating that marker 2 may be 18 marks below or 22 marks above the first marker. There is one student in Figure 1 (next page) who appears to have been awarded over 85% yet with a marker difference of almost 20 marks.

There is no relationship between the difference in marks between the markers (marker 2 – marker 1) and the final mark. Final mark was usually derived from the mean of the two marks.

Figure 1. Bland and Altman (1986) plot of the mean and difference between marker 1 and marker 2.

Comparison was also made between the written comments each marker made about the proposal. A person outside of the Health Program did this analysis. This was to limit the bias introduced in analysis of the comments. This person was obtained through the University and had experience in data coding. The researcher trained the employee and the employee did the data coding for the written comments. The comments for all proposals were coded by the employee. The researcher reviewed 20% of the scripts to assure agreement with the employee doing the data coding. Analysis was made of the comments by looking at the comments made in each of the 24 potential categories as outlined on the feedback form (Appendix B). Comments were analyzed for similarities and differences between the comments within each category. The 24 categories were divided into three subdivisions: (1) those where the comments were the same by both markers, (2) those where the comments were different between the markers and (3) those where some of the comments were the same and some were different within specific categories. Tallies were made under each subdivision. Review of table 2 shows there was generally poor agreement on the comments made by the individual markers. The largest disagreement or partial disagreement occurring in the categories of clarity of title, rationale for the proposal, appropriateness of aims and appropriateness of the research question or testable hypothesis. These categories could be said to be the building blocks of the proposal.

Table 2: Quantitative Comparison of Comments Made By Each Marker

| Comment Category | Number of papers where comments were the same by both makers | Number of papers where the comments were different between the two markers | Number of Papers where there was partial disagreement/ partial agreement |
|--|---|---|---|
| 1. Clarity of title | 32 | 21 | 11 |
| 2. Rationale for proposal | 33 | 19 | 17 |
| 3. Relevance of cited literature | 35 | 17 | 17 |
| 4. Appropriateness of aims: | 28 | 27 | 14 |
| 5. Appropriateness of the research question or testable hypothesis | 30 | 18 | 21 |
| 6. Selection and justification of research design | 45 | 12 | 12 |
| 7. Indication of required number of subjects | 44 | 15 | 10 |
| 8. Indication of method of subject recruitment | 44 | 18 | 7 |
| 9. Organisation of access to possible subjects | 39 | 18 | 12 |
| 10. Indication of inclusion/exclusion criteria of subjects | 43 | 14 | 12 |
| 11. Appropriateness of materials or equipment: | 54 | 9 | 6 |
| 12. Accessibility of equipment (if appropriate) | 48 | 21 | 0 |
| 13. Comprehensive description of procedures | 39 | 20 | 10 |
| 14. Consideration of possible sources of bias | 41 | 13 | 15 |
| 15. Indication of outcome measure(s) including levels of measurement | 48 | 13 | 8 |
| 16. Justification of chosen method(s) of data analysis | 41 | 18 | 10 |
| 17. Consideration of issues of confidentiality and data security | 55 | 8 | 6 |
| 18. Inclusion of information / consent form | 60 | 8 | 1 |
| 19. Indication of safety issues | 56 | 9 | 4 |
| 20. Indication of other ethical issues | 55 | 11 | 3 |
| 21. Description of pilot study | 40 | 17 | 12 |
| 22. Assessment of issues of feasibility, validity and reliability of procedure | 40 | 20 | 9 |
| Indication of planned modifications to procedure in light of above | 50 | 11 | 8 |
| 23. Estimation of resource requirement | 51 | 9 | 9 |
| 24. Reference list | 39 | 25 | 5 |

(n=69 students, this accounts for 3 sets of comments unable to be compared as the comments were not noted within the provided categories)

Observational review of students' age, ethnicity, and gender revealed no significant trends relative to student marks or agreement or differences in comments made by the markers. The small percentage of ethnically diverse students did not allow formal analysis in this area.

DISCUSSION

This study has indicated that there was a wide discrepancy between the classification of the research proposals, the marks given and the comments given by two independent markers. Most often the resultant mark was an average of the two marker's marks and the comments a combination of the two marker's comments. This is consistent with other findings in the literature about double marking, which have also concluded that even after discussion between the two markers the result is most often an average of the two markers on assignments such as essays, case studies, portfolios and oral exams.^{9,12-14} This averaging is a large-scale regression to the mean, making differentiation of students difficult except for those that are exceptionally high or low performers.¹ To date in the literature there is no suggested system for rational decision making when the markers disagree about the quality of a script. In such cases where one marker sees the script as a first class and the second marker sees the script as a third class an averaging of the marks holds little meaning. Researchers allude to the issues of bullying and deference to senior colleagues as ways in which marks are reconciled.¹⁹ In this study issues of deference to senior colleagues did not seem to be present. It has been suggested that if markers belong to the same "community" that chances of deferral to senior colleague and bullying issues would be reduced.²

Double marking was a way to test out the marker's judgement and prevent differences in interpretation.¹⁵ In this study student's marks would have been modified if double marking had been employed. An advantage of using a double marking system is that if a marker makes a significant judgement error in marking a script, the second marker can identify this. In this study some students would have received a different grade than they would have if only one marker had been employed. The downside to this is that for those students who have produced a high quality script and one marker errors on the side of a lower mark, an averaging of the marks results in potentially a lower grade than deserved. If the markers are reliable and produce consistent agreement between marks then the process of double marking may not be needed.⁹ In this case double marking may be a time consuming activity that does not result in any significant value. However, a large discrepancy in marks, with a tendency to regress to the mean, questions the process.

It has been suggested that the reasons that double marking is used should be clearly outlined before the process is employed.¹¹ The greatest discrepancy in marks should occur in the middle ranges, rather than at the high or low end.¹¹ If this is occurring then the double marking process is achieving its aim.¹¹ In this case much discrepancy was found at the high and low end of the scale such as the student in Figure 1 who appears to have been awarded over 85% yet with a marker difference of almost 20 marks. In this situation the double marking could be used as a training tool or a tool to determine the most suitable marking team for the future.

The analysis found generally poor agreement between markers on classification, marks and comments. There are several possible reasons for this. Although all faculty were trained in the classification scheme and marking bands, the potential still existed for differences in interpretations and application of the marking bands. The bands may need to be more specific to help ensure understanding and more consistency in application. The use of a criteria approach to marking could help decrease marking discrepancies and increase the transparency for students and faculty.¹⁵ A combination of the comments made by the two markers has the potential to provide richer feedback to the students, as long as there is no discrepancy between the comments. However, there were many instances in this study where faculty comments opposed each other. Further research is needed to investigate the reasons for discrepancies between markers comments.

Faculty marking the scripts had various levels of experience with mentoring students in the research process, marking student work, conducting their own research and publishing journal articles. These differences in levels of experience may have influenced the expectations of student work. Experience in academia ranged from less than one year to over 20 years. This wide range of differences in experience is sure to have influenced consistency. Additionally, there were 23 faculty members doing the marking. Although the pairs were assigned randomly and resulted in 57 different pairs, the large number of markers increased the potential for differences. The employment of a small core marking team that have certain qualifications may be advisable. Not all faculty may have the knowledge or qualifications to mark the research proposal. The need for a 'community' where open exchange of ideas, values and standards can be shared may be needed.² Further research on how faculty experience is related to the scoring would be beneficial.

The general environment surrounding the marking process may also have influenced the results. In this situation faculty had two weeks to mark anywhere from 4 to 9 scripts based on number of hours worked. This two-week time frame included individual marking, coming to consensus with the other marker and finalizing the marks and comments. Time constraints imposed by full versus part time staff hours and time devoted to other courses may have impacted on the ability and time for staff to adequately reconcile differences. With these constraints, averaging of the marks and combination of the comments may have been the easiest and quickest way to get the job done.

Unlike other studies that were retrospective in nature, this study was prospective and carried out with some structure and controls. It included standard training of the markers and randomisation of the marking pairs as suggested in previous research.⁹ Additional research needs to be conducted to further investigate ways to reconcile differences between markers. If double marking is to be valuable then a structured, effective method of reconciliation of differences between marks in order to get marks based on the level of the script produced is needed. Methods such as negotiation between markers, averaging of marks, employment of a 3rd or 4th marker and a systematic method for weighting differences needs to be further explored. The use of independent and blind moderation also needs further exploration.

RECOMMENDATIONS

Based on this study several recommendations can be made to enhance the transparency of the double marking process. The initial development of marking guidelines with input from the marking team may enhance the breath and depth of the guidelines and increase the commitment to and understanding of the marking guidelines. In conjunction with this the development of a small as marking team as possible with a standard set of qualifications may be useful to enhance consistency between markers. This is in keeping with the idea that markers bring experience of marking other assignments and their area of expertise to the marking of a particular assignment.² Choosing a marking team with similar backgrounds and expertise may increase consistency between markers. Keeping the marking team as consistent as possible from year to year (per assignment) would also increase the consistency in the use of the marking guidelines and understanding of the assignment.

Regular training sessions for staff in the use of the marking guidelines, classification system and marking bands is essential. Training before each marking episode is recommended, for a refresher even when staff has been previously trained. When new staff are added to the marking team, mentoring on how to use the marking guidelines to ensure consistency in application of standards would be beneficial. Peer review marking, shadow marking, rank ordering marks and detailed discussion of the use of marking schemes should also be considered.¹¹

Quality of the marking guidelines must also be considered. Some authors have suggested that specific marking criteria prevent markers from bringing a completely subjective assessment to the assignment.^{15,20,21} Other authors have suggested that specific criteria are open to interpretation and the background of the reader.² Guidelines that are specific yet leave room for valuing the wholeness of the product with the inclusion of point values and or percentage of weighting may be useful for consistent application among markers. The development of guidelines on how to reconcile differences and arrive at the final mark needs consideration in addition to set standards for when 3rd markers are needed. For example if initial marks are a certain point value or class value apart 3rd marks could be / should be employed. In order for 3rd marks to be easily accepted by staff and not seen as a negative, reinforcement that differences in opinion does not constitute poor marking ability but highlights areas where additional guidance is needed. To this end it is important to ensure that there is adequate time for marking and meetings between markers to reconcile differences.

CONCLUSION

This study has looked at double marking of a research proposal in a health course. The outcomes are similar to previous studies that have looked at double marking of various other types of written assignments. Further exploration of the double marking process and other marking processes that results in a transparent system for faculty and students is needed. Exploration of how faculty's background and experience influence marks given would be beneficial. Further exploration of the influence of different types of marking rubrics and marking guidelines would also add to the body of knowledge in this area and could result recommendations for more transparent marking systems.

REFERENCES

1. White R. 2005. Double Marking Versus Monitoring of Examinations. The Higher Education Academy, Centre for Philosophical and Religious Studies (WWW) Available from: http://prs.heacademy.ac.uk/documents/discussion_articles/double_marking_versus_monitoring_of_examinations.html Accessed 3 May 2007.
2. Bennett J. Second-marking and the academic interpretative community: ensuring reliability, consistency and objectivity between markers. *Investigations in Teaching and Learning*. 2005; 4(1): 80-86.
3. Brooks V. Double Marking Revisited. *British Journal of Educational Studies*. 2004; 52(1): 129-147.
4. Rowntree D. *Assessing students: how shall we know them?* Kogan Page, London. 1987
5. McVey P. A comparison of reliability: objective and conventional examinations in electrical engineering. *International Journal of Electrical Engineering Education*. 1978; 15: 109-114.
6. Weingarten M, Polliack M, Tabenkin H. Variations among examiners in family medicine residency board oral examinations. *Medical Education*. 2000; 34: 13-17.
7. Pitts J, Cole C, Thomas P. Educational portfolios in the assessment of general practice trainers: reliability of assessors. *Medical Education*. 1999; 33: 515-520.
8. Dennis I, Newstead S, Wright D. A new approach to exploring biases in education assessment. *British Journal of Psychology*. 1996; 87: 515-534.
9. Smith B, Sinclair H, Simpson J, Van Teijlingen E, Bond C, Taylor R. What is the role of double-marking? Evidence from an undergraduate medical course. *Education for Primary Care*. 2002; 13: 497-503.
10. Williams R, Sanford J, Stratford P, Newman A. Grading written essays- a reliability study. *Physical Therapy*. 1991; 71(9): 679-686.
11. Cannings R, Hawthorne K, Houston H. Putting double marking to the test: a framework to assess if it is worth the trouble. *Medical Education*. 2005; 39: 299-308.
12. Sparks R, Ballanyne R. Quality control methods in large-scale assessment procedures using 'double marking' or 'partial double marking'. *Quality Control and Applied Statistics*. 1997; 42(1): 45-48.
13. Partington J. Double-marking students' work. *Assessment and Evaluation in Higher Education*. 1994; 19(1): 57-60.
14. Newstead S, Dennis I. Examiners examined: the reliability of exam marking in psychology. *The Psychologist: Bulletin of the British Psychological Society*. 1994; 7: 216-219.
15. Hand L, Clewes D. Marking the differences: an investigation of the criteria used for assessing undergraduate dissertations in a business school. *Assessment and Evaluation in Higher Education*. 2000; 25(1): 5-21.
16. Altman D. *Practical Statistics for Medical Research*. Chapman and Hall. London. 1991.
17. Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; i: 307-310.
18. StataCorp. Stata Statistical Software: Release 9.0 College Station, Texas: Stata Corporation. 2005.
19. Ecclestone K. "I know a 2.1 when I see it": understanding criteria for degree classifications in franchised university programmes. *Journal of Further Higher Education*. 2001; 25(3): 301-313.
20. Webster F, Pepper D, Jenkins A. Assessing the undergraduate dissertations. *Assessment and Evaluation in Higher Education*. 2000; 25(1): 71-80.
21. Sandhu S. The usage of statement banks as a means of assessing large groups. *Investigations In University Teaching and Learning*. 2004; 2(1): 49-55.

APPENDIX A
General Classification and Band Definitions¹

| Classification | Band | Definition | Descriptor |
|-----------------------|----------------|-------------------|--|
| 1st | 70-100% | Excellent | Outstanding work which displays consistent originality; evidence of very thorough and mature intellectual engagement with complex themes and discussion. Authoritative grasp of relevant concepts, knowledge and information appropriate to the subject discipline; evidence of extensive reading of relevant books and journal articles and robust engagement with primary sources; ability to critically analyse and evaluate evidence and argument at a high level; ability to critically appraise methodology of evidence and develop arguments logically and clearly at a high level; demonstrates clear awareness of his or her own value judgements and assumptions; indications of originality in application of ideas, in synthesis of material; very high level of competence in correct and clear English usage with very few or no imprecise statements. |
| 2.1 | 60-69% | Very Good | Sound level of understanding based on a competent grasp of relevant concepts, knowledge and information appropriate to subject discipline, evidence of reading a good range of relevant books and journal articles with attention to primary sources; sound ability to critically analyse and evaluate evidence and argument; sound ability to critically appraise methodology of evidence and develop arguments logically and clearly; able to demonstrate clear awareness of own value judgements and assumptions; displays skill and some originality in interpreting complex material; high level of competence in organisation of material; correct English usage with very few imprecise statements. |
| 2.2 | 50-59% | Good | Demonstrates a coherent response to requirements of the assessment task; clear expression of ideas; accurate restatement of relevant source material; evidence of reading key texts and journal articles; shows some ability to evaluate evidence and argument and expose value judgements. Characterised by assimilation rather than integration; some understanding of the strengths and weaknesses of the methodology being used; draws recognisable and relevant conclusions; good organisation of material; correct English usage with few imprecise statements. |
| 3rd | 40-49% | Acceptable | Recognisable awareness of requirements of assessment task; evidence of limited knowledge of relevant source material and limited use of primary sources; limited attempt to organise a response to themes; some attempt to draw relevant conclusions; correct English usage; some imprecise statements. Clear but still limited appreciation of links between practice experience/knowledge and relevant concepts/theory; some signs of ability to apply theoretical issues to practice. |
| Fail | 37-39% | Fail | Clear fail, not eligible for compensation. Work deficient in most respects, revealing insufficient grasp of material and/or poor organisation. Inability to identify and address task required. |

¹ General Classification and Band Definitions: A student whose score is 70% or greater is said to have a 1st classification. A student who obtains a score between 60-69% is said to have a 2.1 classification and so on with each classification.

APPENDIX B
Research Proposal
Data Collection Form

Student Number _____

Staff Number _____

| CLASS | | MARKING BANDS |
|--------------|------------|----------------------|
| 1st | Excellent | 70-100 |
| 2.1 | Very Good | 60-69 |
| 2.2 | Good | 50-59 |
| 3rd | Acceptable | 40-49 |
| FAIL | Fail | 0-39 |

Marks by Independent Staff Member Before Comparison

1. Class: _____
2. Specific Mark _____
3. Specific Written Comments: See attached sheet.

Please make comments on the following items: adjust spacing as needed to insert comments

Title of Proposal:

1. Clarity of title:
2. Rationale for proposal:
3. Relevance of cited literature:
4. Appropriateness of aims, research question or testable hypothesis:
5. Selection and justification of research design:
6. Indication of required number of subjects:
7. Indication of method of subject recruitment:
8. Organisation of access to possible subjects:
9. Indication of inclusion/exclusion criteria of subjects:
10. Appropriateness of materials or equipment:
11. Accessibility of equipment (if appropriate)
12. Comprehensive description of procedures:
13. Consideration of possible sources of bias:
14. Indication of outcome measure(s) including levels of measurement:
15. Justification of chosen method(s) of data analysis:
16. Consideration of issues of confidentiality and data security:
17. Inclusion of information / consent form:
18. Indication of safety issues:
19. Indication of other ethical issues:
20. Description of pilot study:
21. Assessment of issues of feasibility, validity and reliability of procedure:
22. Indication of planned modifications to procure in light of above:
23. Estimation of resource requirement:
24. Reference list: