

**SHIFT FROM ORAL TO WRITTEN EXAMS OF PHARMACOLOGY
AND INTER-RATER DIFFERENCES**

Slobodan M. Jankovic, Snezana V. Jankovic

Pharmacology department, Medical Faculty, University of Kragujevac,
Kragujevac, Serbia & Montenegro

Summary

Oral exams are a subjective method of evaluating pharmacology students' knowledge and depend on the rater's performance and attitude. In our study we have investigated the effects of introduction of written exam with dual assessment at undergraduate pharmacology and toxicology course, on inter-rater variability in grading. The students are writing an essay during the exam, which is then assessed by two independent raters. The two periods were compared, before and after introduction of the written exams with dual assessment. In the first period, two raters were giving significantly different grades, while in the second differences were not significant any more. The total average grades did not change, but passing rate of the exam decreased for one-third. The results of our study suggest that written exams with dual assessment contribute to objectivity of assessment of students' knowledge in pharmacology and toxicology.

Key Words: pharmacology, assessment, undergraduate, inter-rater reliability.

Address all correspondence to:
Prof. Dr Slobodan M. Jankovic
Medical Faculty, University of Kragujevac
Ul. Svetozara Markovica 69
34000 Kragujevac, Serbia, Yugoslavia.
tel. ++381 64 1168103
Fax. ++381 34 370073
e-mail: slobnera@EUnet.yu

INTRODUCTION

Small differences in number grades of medical students are statistically meaningful, giving to educators an opportunity to use remedial education and improve final outcome¹. The number grades of medical students are also strongly correlated with future success in academic career². Valid and fair rating of students' performance on examinations is of crucial importance for success of medical education.

However, oral exams are a subjective method of evaluating students' knowledge and depend on the rater's performance and attitude³. Written exams with double rating should be used for objective student assessment.

For many years, after completing undergraduate pharmacology and toxicology course at Medical Faculty, University of Kragujevac, students were sitting oral exams. However, significant differences in rating students were noted between the two available raters; three years ago oral exams were replaced by written exams, with double rating of students' exam essays. The aim of this study was to compare inter-rater variability before and after introducing written exams.

MATERIAL AND METHODS

The pharmacology and toxicology undergraduate course at Medical faculty, University of Kragujevac, is a two-semester, fourth-year course, teaching students to both basic and clinical pharmacology and toxicology. From year 2000 onwards, two raters were available for this course, a professor and a docent (lecturer). From 2000 to 2002, both raters were examining students orally and independently, while from 2002 to 2004, students were writing exam essay, and both raters were rating the essay, changing the order of rating from one to another exam term.

From 2000 to 2002, rater one (professor) assessed 98 students, and rater 2 (docent) 88 students. From 2002 to 2004, 595 students wrote the exam essay, and were rated by both raters.

The exam grades were taken from the official exam book; methods of descriptive statistics were used for summarizing the data.

RESULTS

In the period 2000 – 2002, rater 1 assessed 98 students, by oral exam. Sixty-three students (64.29%) passed the exam, and were rated on the scale from 6 to 10, as following: 15 students were rated 6, 29 were rated 7, 14 were rated 8, 4 were rated 9 and one was rated 10 (Figure 1). The average grade rated by the rater 1 was 7.16 ± 0.91 (SD).

In the period 2000 – 2002, rater 2 assessed 88 students, by oral exam. Sixty-two students (70.45%) passed the exam, and were rated on the scale from 6 to 10, as following: 5 students were rated 6, 9 were rated 7, 9 were rated 8, 24 were rated 9 and 15 were rated 10 (Figure 1). The average grade rated by the rater 2 was 8.56 ± 1.24 (SD).

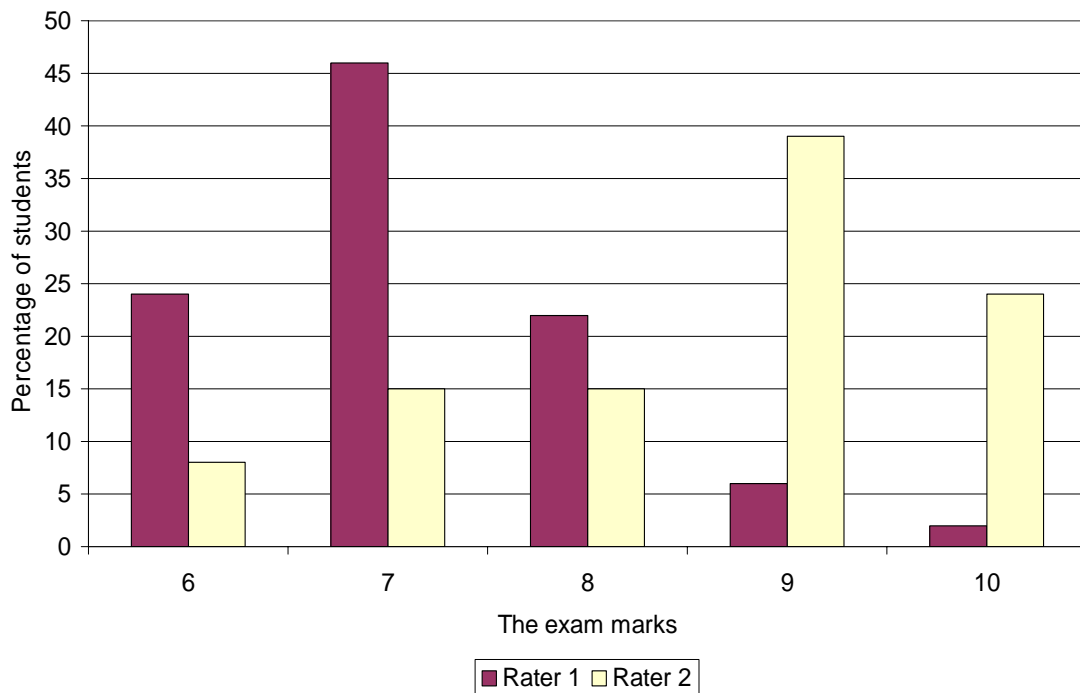


Figure 1. Frequency distribution of grades from oral exams for the period 2000-2002.

The chi-square test for association (contingency table) showed highly significant relationship between the type of the rater and frequency distribution of exam marks (chi-square test value 21.399, $df = 4$, $p < 0.001$).

In the period 2002-2004, the raters assessed 595 students, by written exam. Each student was assessed by both raters. Two hundred and thirty-three students (39.16%) passed the exam, and were rated on the scale from 6 to 10, as following: (1) rater 1: 56 students were rated 6, 78 were rated 7, 68 were rated 8, 26 were rated 9 and 5 were rated 10; (2) rater 2: 61 student was rated 6, 74 were rated 7, 63 were rated 8, 28 were rated 9 and 7 were rated 10 (Figure 2). The average grade rated by the rater 1 was 7.34 ± 1.03 (SD), while that of the rater 2 was 7.34 ± 1.08 (SD).

The chi-square test for association (contingency table) showed absence of relationship between the type of the rater and frequency distribution of exam marks (chi-square test value 0.901, $df = 4$, $p > 0.05$).

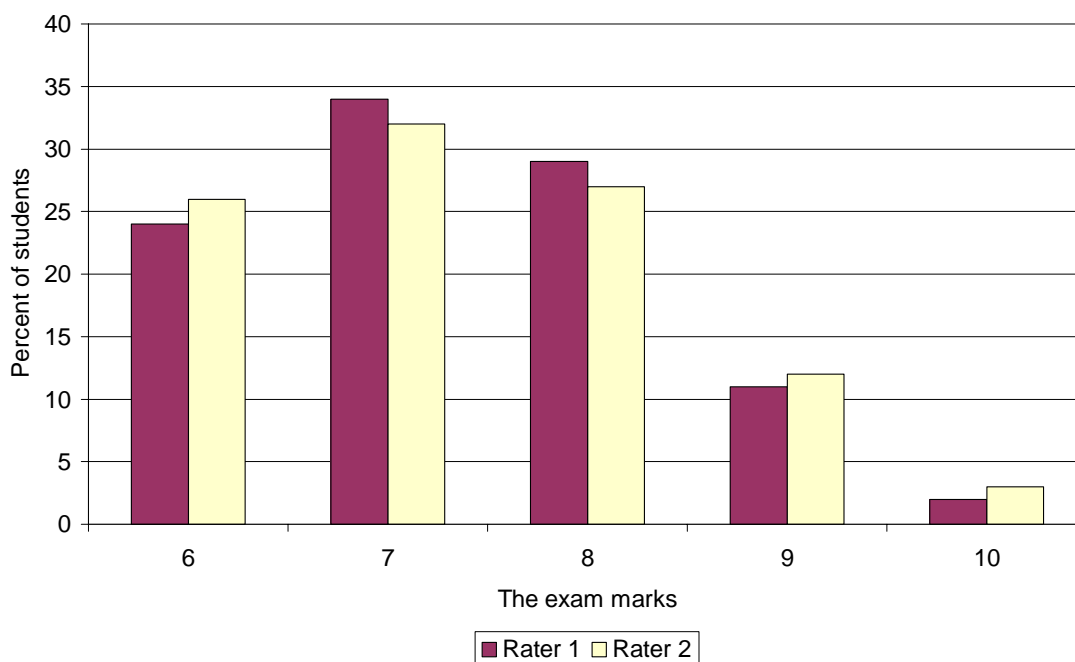


Figure 2. Frequency distribution of grades from written exams for the period 2002-2004

DISCUSSION

Our results clearly show significant decrease of rating variability between the two raters; the comparison between Figures 1 and 2 is very illustrative of this change. The higher inter-rater reliability was result of introduction of written exam with double rater assessment. Although some studies showed satisfactory inter-rater reliability with oral exams⁴, written exams contribute to objectivity of assessment, and dual assessment of the same students' essays does it even further. Each of the two raters is aware of another rater's attitudes and opinions, and tries to compromise with its own criteria for rating.

If traditional oral exams in pharmacology and toxicology courses are kept, over time upward trends in grades could be observed⁵. Raters are subject to multiple influences from a society, and sometimes it is difficult to keep the same assessment criteria over years. However, if dual assessment is put into the practice, the influences are much easier to resist. Besides, written exams leave durable written proof of what was really happening at the exams.

Introduction of written exams with dual assessment did not decrease the average grade, but percentage of students who passed the exams decreased for one third. It could seem as an drawback of this examination method, but more objective assessment contributes to higher knowledge and skills levels of graduated medical students; it was shown that more rigorous approach during undergraduate pharmacology courses led to higher scores of students at state examinations⁶.

REFERENCES

1. Gonnella JS, Erdmann JB, Hojat M. An empirical study of the predictive validity of number grades in medical school using 3 decades of longitudinal data: implications for a grading system. *Med Educ* 2004; 38: 425-34.
2. Kuzman T, Bergovec M, Rojnic M. Were today's professors good students? Case study of the Zagreb University School of Medicine. *Croat Med J* 2004; 45: 54-8.
3. Golubic R, Golubic K. What do grades in clinical subjects depend on? Case study of the Zagreb University School of Medicine. *Croat Med J* 2004; 45: 67-71.
4. Allison R, Katona C. Audit of oral examinations in psychiatry. *Med Teach* 1992; 14: 383-9.
5. Bergovec M, Kuzman T, Rojnic M, Makovic A. Is there grade inflation at medical schools? Case study of the Zagreb University School of Medicine. *Croat Med J* 2003; 44: 92-7.
6. DelliPizzi A. A return to the past: a student perspective on medical school pharmacology. *J Clin Pharmacol* 2000; 40: 39-43