Are we consistent? The effects of digitized exams grading

Migut, Gosia; Wiersma, Ruben

Citation (APA)
Migut, G., & Wiersma, R. (2020). Are we consistent? The effects of digitized exams grading. In *SIGCSE 2020 - Proceedings of the 51st ACM Technical Symposium on Computer Science Education* Association for Computing Machinery (ACM). https://doi.org/10.1145/3328778.3372630

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Are We Consistent? The Effects of Digitized Exams Grading

Gosia Migut
Delft University of
Technology
The Netherlands
m.a.migut@tudelft.nl

Ruben Wiersma
Delft University of
Technology
The Netherlands
rubenwiersma@gmail.com

ABSTRACT

Many universities digitize exams or the process of grading the exams. This potentially allows for faster grading, is less labor intensive and less error-prone. But are the grades produced by online grading consistent with how we grade on paper? In this paper we present preliminary results of the comparison between scores given by grading online and grading on paper.

KEYWORDS

digitized exams, online grading, higher education

1 INTRODUCTION

Evaluating exam results with credibility is time consuming and error prone. Unsurprisingly, there is a growing push to digitize paper exams and grade them online, removing the bound to a physical location while improving quality as well as grading speed [1] .

The common procedure is to digitize answer sheets and allow graders to access them online. Several commercial interfaces are available and the users of them claim to improve their efficiency and transparency, e.g. [1]. There is however no evidence on whether the online grading strategy has influence on the grade distribution.

Online grading system typically use a different workflow from the traditional, paper based method. Online you are allowed to assign a pre-fixed set of feedback options with the corresponding question grade. In the paper method, even if a strict rubric is used, the final grade assignment is done not by choice of one of the pre-fixed sets, but can be done more subjectively.

In this paper we investigate whether the distribution of grades produced by online and paper grading correspond. Our hypothesis is that because in the online grading you have to select one feedback with pre-fixed set, the grading will be done more strict. In the paper grading when only grades are awarded there is more freedom in giving a slightly higher or lower grade based on subjective reasons.
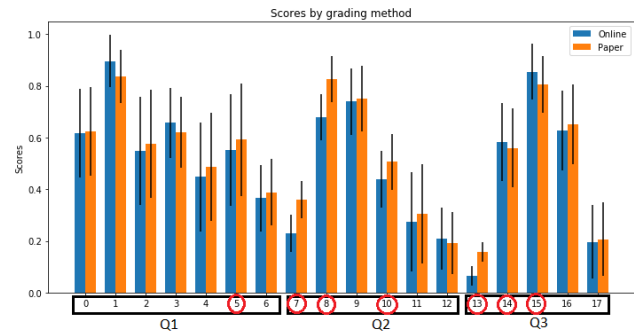
Figure 1: The mean and standard deviations of sub-question grades for online and paper grading. The red circles denote sub-questions that are graded significantly different.

2 GRADING OF DIGITIZED EXAMS

At Delft University of Technology, an online system called Zesje[1] is gaining popularity. After digitizing the paper answer sheets, grading is done online. Per question, the grader chooses how many points to award. Each possible amount of points is accompanied by a pre-fixed feedback. This differs from the paper grading where no feedback for the student is provided at all, only points are assigned.

We did preliminary experiments with 6 graders and 3 large questions with sub-questions per exam. Each question is graded online and on paper. Each grader always graded the same question in each exam and graded half of the exams online and half on paper. Each exam is thus graded both online and on paper by 2 graders. The course was for the second year BSc course, we graded 115 exams.

A t-test shows a significant difference ($p = 0.003$) between online ($\mu = 6.06$) and paper grading ($\mu = 6.27$). For 5 sub-questions the mean is significantly higher for the paper grading, for 2 sub-questions the mean is significantly higher for online grading, see Fig 1. Our setup was inconclusive if there was a difference between the time spent on grading the exams.

3 DISCUSSION

Our preliminary results show a small but significant difference between online grading and paper grading confirming our hypothesis. The next step is collecting more data, evaluating question types, student response, and investigating time spent on grading.

REFERENCES

[1] A. Bloomfield. Evolution of a digital paper exam grading system. *2010 IEEE Frontiers in Education Conference (FIE)*, pages T1G–1–T1G–6, 2010.

---

[1] https://quantumtinkerer.tudelft.nl/blog/zesje/