Negative marking, guessing and academic performance

Ghazala Azmat (Sciences Po), Maia Guell (University of Edinburgh) and Stefania Simion (University of Bristol)

Thursday 2 September, 15:00-16:30 BST

Little is known about the consequences of penalizing bad performance (or mistakes). In the context of education, examiners can adopt a negative marking scoring system as a way to discourage students from guessing and to increase test reliability and validity. One common form of assessment that facilitates this type of system are multiple-choice exams. Multiple-choice exams that use a negative scoring system often penalize students for wrong answers, giving no marks for blank answers and rewarding correct answers. However, an important concern of this system is that it might reduce the performance of those who are relatively less confident in their likelihood of answering correctly or are more risk averse. In turn, inducing a trade-off in efficiency and equality.

In this paper, we use a quasi-experimental design at a large university in the UK to explore the performance impact of using a negative marking scoring system. In particular, we explore a reform in 2014, whereby change in the format of summative multiple-choice tests - from 100% negative marking to 50% negative marking - in two core economics undergraduate courses. This natural experiment accounts for any framing biases inherent in controlled experiments and provides a real-world measure of the effect of negative marking.

We aim to contribute to the emergent literature on the effects of negative marking in multiple-choice tests (Atwater and Saygin, 2021; Iriberri and Rey-Biel, 2021; Coffman and Klinowski 2020; Funk and Perrone, 2017) by focusing not only on the overall effects, but also on a set of heterogenous ones. First, given the existing evidence that women are more risk averse than men (Eckel and Grossman, 2008; Croson and Gneezy, 2009) and less confident (Barber and Odean, 2001), we analyse whether female students are more likely to be discriminated against by negative marking. Then we explore whether the effects of negative marking are different among students with different levels of ability, as well as fields of study. Finally, given that the two courses that we analyse apply different penalties for incorrect answers, we also plan to study whether the results are sensitive to the magnitude of the imposed penalty, a new dimension not explored before in the literature.

We use a rich individual level dataset, covering 4 cohorts of students, two exposed to 100% negatively marked exams (2012-2013) and two exposed to 50% negatively marked exams (2014-2015). Students take multiple choice exams as mid-terms through the year, as well as for part of their final exam. We measure their test performance, including the answers (or blanks) given to each question in each test.

In our analysis, we employ a within-individual regression analysis to explore the variation across two parts of the exams - part 1, which was negatively marked before 2014, but not negatively marked post 2014 and part 2, which was negatively marked throughout the entire period under analysis. Our regression allows us to identify changes in performance overall in the part 1, as well as any potential spillover onto part 2.

Our preliminary findings suggest that scrapping negative marking increases the grades, encouraging students to leave less blanks. However, it also increases the proportion of incorrect answers, to the extent that the overall number of correct answers under the new regime does not increase. The effects are similar for students with different socio-demographic characteristics and ability. The implications of our preliminary results imply that, despite students getting higher marks when there is no negative marking, their learning has not improved. Thus, introducing negative marking in multiple-choice exams seems to be efficient in improving knowledge, by reducing guessing, with no strong heterogenous effects. In ongoing work, we aim to analyse more drastic changes in negative marking, in terms of the applied penalty, to understand the importance of the size of the imposed penalty.