



Luciana Maia Campos Machado,
Claudia Emiko Yoshinaga
Brazil

New Educational Review

Self-assessment Accuracy, Overconfidence and Student Performance

DOI: 10.15804/tner.2018.53.3.07

Abstract

We have analyzed results of the Brazilian National Student Performance exam, applied by the Ministry of Education, to verify overconfidence in students. Looking at 549,487 student-level observations, we estimated an overconfidence score, comparing the perceived performance with the actual grade in two different parts of the exam. Ordered logit models suggest that overconfidence in Brazilian students is positively correlated with income and that overconfident students took less time to complete the exam. Contrary to previous studies conducted in other countries, male students presented lower overconfidence in our sample. Moreover, performance was inversely related to the overconfidence score, indicating that students with better performance “know more about what they do not know”, as already widely discussed in the education literature.

Keywords: *overconfidence, student performance, self-assessment accuracy.*

Introduction

People often believe they are in a better situation than they really are (Lichtenstein and Fischhoff, 1977). Several studies analyzed the tendency of people to be overconfident. The most common expression of this bias is considered to be the better-than-average effect. Authors like Svenson (1981) found that the majority of the population considers themselves to be a better driver than average. Another manifestation of the better-than-average effect is unrealistic optimism, in which

people believe they will be more successful in their work (Weinstein, 1980), marriage (Baker and Emery, 1993) and other positive life events. Optimistic biases are a common feature when we analyze people's predictions and expectations (Armor and Taylor, 2002). Overconfidence is also thought to manifest itself as self-attribution bias, when people regard success as a consequence of their skills, but they exempt themselves from blame when failures occur (Taylor and Brown, 1988).

In the education context, prior studies often conclude that students overestimate their performance on exams and this overconfidence increases as their real performance (grades) decrease.

Kennedy et al. (2002) were some of the first to identify that when questioned about their performance after an exam, students tend to overestimate their performance. The authors attribute this effect to cognitive abilities and suggest that smarter individuals tend to underestimate their performance, while those with less memory and intellectual ability overestimate their outcome because they do not "know what they [effectively] know."

Miller and Geraci (2011) also suggest that people can be overconfident because they have poorer abilities. In their work, students with poor performance presented higher overconfidence when compared to high-performance students. According to the authors, probably because poor students have some awareness of their lack of metacognitive knowledge. Koku and Qureshi (2004) and Langendyk (2006) found strong evidence that high performance students are able to better discriminate their real knowledge and usually present lower levels of overconfidence. However, Clayson (2005), analyzing determinants of overconfidence, found no evidence that overconfidence stems from a lack of cognitive competence. In fact, the author believes past experiences and student expectations can be determinant in this behavior.

Bengtsson et al. (2005) and Pallier (2003) found that exam behavior is gender-specific when analyzing data from tests at Stockholm University: male students were more overconfident than women. Nowell and Alston (2007) and Dahlbom et al. (2011) found very similar results when analyzing grades expectations of economics students and 14-year-old high school students. Men were overconfident about their grades, whereas women were consistently underconfident.

Relying on an official database containing 549,487 observations at student level, made available by the Ministry of Education after the ENADE (National Student Performance Exam) application, we estimated an overconfidence score by comparing perceived performance with the actual grade in two different parts of the exam, being "General Knowledge", content beyond the field of the student's training and "Specific Knowledge", specific contents of each undergraduate program.

To the best of our knowledge, this is the first study to use this database to measure students' overconfidence and relate it to performance, gender and socioeconomic conditions.

Research Methodology

Data and Empirical Strategy: ENADE Microdata

Our sample consists of 549,487 student-level observations of the ENADE (National Student Performance Exam) of 2015 (INEP, 2017). Our database contains information on 1,758 Brazilian higher education institutions from 28 different states and 5,570 municipalities of the country. Of the 549,487 students in the sample, 58% are female and 42% are male.

The ENADE is part of SINAES (National System for the Evaluation of Higher Education) and aims to evaluate the quality of higher education courses in Brazil.

Students' performance is evaluated by a test applied to students of various undergraduate programs every three years. ENADE aims to assess students' skills and abilities from specific contents of each undergraduate program (Specific Knowledge) and content beyond the field of training (General Knowledge).

Data obtained from students' answers are publicly available, without nominal identification of students. The ENADE final grades vary from 1 to 5 and are only publicly available as a collective outcome of the institution program. Individual results are disclosed exclusively to each student in a private environment. A collective result inferior to 3 ends up exposing the higher education institution to media and competitors, leading to negative reputation.

Besides the test, students are also given a questionnaire about their socioeconomic profiles, how they evaluate the program (and institution) in which they are enrolled and their perception of performance in the exam. Students' evaluation through the questionnaire and collective results obtained by them in the ENADE are used as quality measures in the composition of a Preliminary Course Concept (CPC).

The Preliminary Course Concept (CPC) is an indicator of quality that evaluates undergraduate programs in Brazil. Its calculation and official release occur in the year following ENADE results, based on the evaluation of i) students' performance; ii) value added by the learning process and iii) inputs related to supply conditions (teaching staff, infrastructure and didactic-pedagogical resources), according to technical guidance approved by the National Commission for the Evaluation of Higher Education (CONAES). Courses that do not have at least two graduating

students do not have their calculated CPC, remaining as a “no concept course” (INEP, 2018).

A selected sample of first and senior year students are submitted to the exam every three years. For selected students who are graduating, participation in the ENADE is mandatory and an indispensable condition for the issuance of the school record. Unselected students can also take the test as volunteers.

In 2015, students of the programs that confer a bachelor's degree in Business Administration, Public Administration, Accounting, Economics, Social Communication - Journalism, Social Communication - Advertising and Propaganda, Design, Law, Psychology, International Relations, Executive Secretariat, Theology and Tourism, and programs that confer a diploma of Technologist in Foreign Trade, Interior Design, Fashion Design, Graphic Design, Gastronomy, Commercial Management, Quality Management, Human Resources Management, Financial Management, Public Management, Logistics, Marketing and Management Processes were submitted to the exam.

The exam is divided into questions related to *General Knowledge*, which represent 25% of the final grade, and *Specific Knowledge* of each course, representing 75% of the final grade. At the end of the exam, students fill in a questionnaire about socioeconomic information and perception of the test and course in which they are enrolled. One of the questions addressed in this questionnaire is the degree of difficulty experienced when performing each of the parts of the test (General Knowledge and Specific Knowledge), namely: very easy, easy, medium, difficult and very difficult.

Overconfidence Measure Construction

Although recent education theories attribute low student performance to the finding that overconfident students strive less and tend to perform poorly, there are two explanations about the inverse relationship between overconfidence and student performance.

The first explanation is that students with lower cognitive abilities find it difficult to recognize their real level of ability. The second justification for this miscalibration behavior is that this is a systematic effect, related to past experiences and expectations. Clayson (2005) examined these two hypotheses and found no evidence that the relationship is due to lack of cognitive ability.

In order to minimize the possibility that overconfidence found is related only to students' cognitive perception of their actual performance, we analyzed the scores obtained in both parts of the exam (General Knowledge and Specific Knowledge), considering the distribution of the difficulty perception of the exam. We divided

the grades thresholds according to students' assessment distribution, in order to reduce any possible bias arising from the actual difficulty of the questions proposed in the test, i.e., we tried to minimize any effects of bad calibration because of the high difficulty of the proposed questions in the exam.

The next intervals represent, according to the difficulty described by students, the corresponding performances, as shown in Figure 1. We assigned grade intervals according to the students' perception of the exam. The percentages refer to the number of students who considered the test very difficult, difficult, medium, easy and very easy. For example: if 2% of the students found General Knowledge questions "very easy", the correspondent interval that represents the adequate performance – being excellent – is between 85 and 99.2 points (maximum grade).

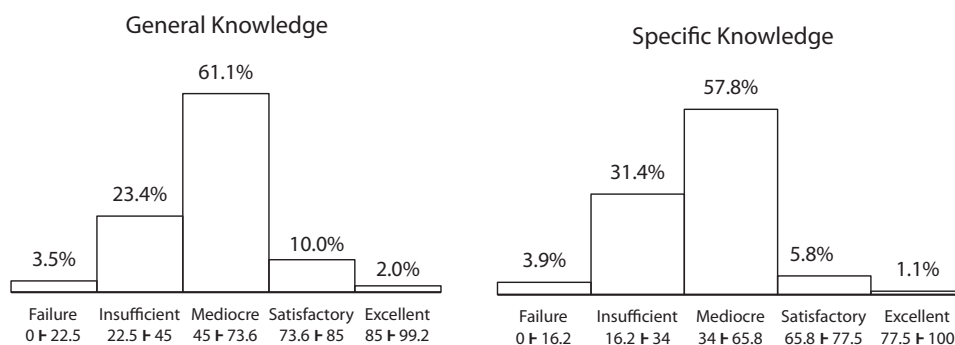


Figure 1. Students' corresponding performance, considering grade distribution

Lichtenstein and Fischhoff (1977) claim that a judgment is perfectly calibrated if, in the long run, probabilities assigned by one individual are equivalent to those actually performed. That is, overconfidence happens when the difference between the mean of the probability responses and the overall proportion are shown by a positive difference. At the other extreme, when the difference becomes negative, underconfident behavior is considered.

Based on the same assumption, we then compared the perceived difficulty interval with each student's effective performance interval by assigning them a score, based on how far the perceived performance was from actual performance, ranging from -4 to 4 (where -4 is the lowest degree of overconfidence and +4 is the highest one). If a person has an overconfidence score of 0, it means that they made the correct assessment of their own performance in the exam, considering test difficulty. The results are presented in Table 1.

Table 1. Overconfidence score construction

Student Perception of Exam Complexity					
Student performance <i>*based on grades distribution</i>	Very Easy	Easy	Medium	Difficult	Very Difficult
Excellent	0	-1	-2	-3	-4
Satisfactory	1	0	-1	-2	-3
Mediocre	2	1	0	-1	-2
Insufficient	3	2	1	0	-1
Failure	4	3	2	1	0

Table 1 shows our overconfidence score construction. The students’ perception of exam complexity corresponds to the classification assigned by the students to the exam difficulty, in a scale of very easy to very difficult. Student performance corresponds to the effective classification every student obtained in the exam, according to grade distribution. This distribution was constructed based on the respective percentile considering the difficulty intervals related to the students.

Figure 2 presents the overconfidence score distribution (ranging from -4 to 4), separately for the two parts of the exam, general and specific knowledge.

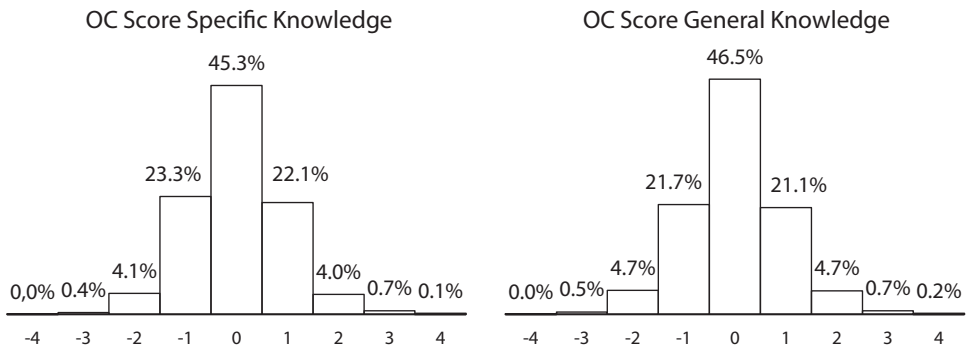


Figure 2. Overconfidence score distribution

First analysis suggests that the overconfidence score follows a normal distribution, with most students classified as having overconfidence 0. Still, there is a fair number of students presenting underconfidence or overconfidence, whose characteristics related to this possible bias can be addressed.

Research Results

We estimated an ordered logit model, where the probability of observing *Overconfidence* corresponds to the probability that the following linear function is within the range of cutpoints estimated for the outcome:

$$\Pr(\text{overconfidence}_j = i) = \Pr(K_{i-1} < B_1 \text{Performance} + B_2 \text{Gender} + B_3 \text{High Income} + B_4 \ln \text{Age} + B_5 \text{Exam Duration} + X + u_j \leq k_i)$$

where *Overconfidence* is our measure of students' overconfidence ranging from -4 to 4, *Performance* is the official grade students have in the "general knowledge" and "specific knowledge" parts of the exam; *Gender* is a dummy that assumes the value of one if the student is male; *High Income* is a dummy that assumes the value of one if the student has a monthly income higher than 4.5 minimum wages (above R\$ 3.258,01); *Age* is the log of the student's age; *Exam Duration* corresponds to the time it took the student to complete the test; and *X* is a vector of control variables that includes geographical region and course.

Table 2. Ordered Logit regression results

	General Knowledge	Specific Knowledge
Performance	-0.0908*** (-182.63)	-0.0966*** (-132.76)
Gender (Male = 1)	-0.161*** (-17.52)	-0.151*** (-18.18)
High Income	0.235*** (19.67)	0.151*** (14.29)
Age (ln)	-0.569*** (-18.23)	-0.325*** (-12.56)
Exam Duration	-0.218*** (-25.52)	-0.251*** (-28.63)
<i>Region</i>	Yes	Yes
<i>Course</i>	Yes	Yes
N	425,160	425,350
Pseudo R ²	0.1668	0.175

Table 2 presents ordered logit regression results for the parts of the exam corresponding to general and specific knowledge. Standard errors are clustered by

educational institution. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Results suggest that, as found by Miller and Geraci (2011) and Koku and Qureshi (2004), student performance is inversely correlated with overconfidence. One of the explanations for this behavior is that best performing students “know more about what they know” (Lichtenstein and Fischhoff, 1977).

We also found that the students who finished the exam faster tended to have higher overconfidence. This behavior suggests that, judging the exam “easy” in relation to their knowledge, the students spent less time answering the questions. They were less careful during the exam.

All these results are statistically significant and present in both general and specific knowledge parts of the exam.

Discussion

The question of “knowing about what is known” is addressed in the work of Lichtenstein and Fischhoff (1977): “Do those who know more also know more about how much they know?” “The effectiveness of self-assessment is divided by the authors into components of overconfidence, underconfidence, calibration and resolution. They found that people who have greater knowledge about their actual performance, i.e., greater calibration, tend to outperform those who exhibit over – or underperformance.

Unlike Nowell and Alston (2007) and Dahlbom et al. (2011), male students in Brazil presented less overconfidence than female students. In literature from other areas, such as finance (Barber and Odean, 2001; Deaves, Lüders and Luo, 2008; Beckmann, and Menkhoff, 2008, among others), women appear to be less overconfident than men. Our findings suggest that, within knowledge-related environments, there is room for further study on gender issues and how they interact with other social and education variables.

The fact that the student has a high income is positively related to overconfidence. Our high-income variable, when analyzed, was positively correlated with international experience and study in private schools (which in Brazil are known to offer higher quality education). We chose not to add these variables to the main regression because of their correlation with high income, which would bring endogeneity to our model. However, this relation may be important in reflecting on overconfidence behavior. Do students who see themselves in better social conditions in an emerging country tend to consider themselves above average?

Age is inversely correlated with overconfidence, so older students were more able to identify their real performance. This behavior suggests that, over the years and with greater maturity, students will acquire greater autonomy and be able to evaluate themselves more efficiently.

Conclusions

This study aims to verify if the previous findings of education literature, which relate overconfidence bias to performance and students' personal characteristics, apply in the Brazilian context.

Using a large sample of student-level observations of the ENADE (National Student Performance Exam) of 2015, released by the Brazilian Minister of Education, we compared the perceived difficulty interval with each student's effective performance, by assigning them a score based on how far their perceived performance was from actual performance.

As the Ministry of Education regulates undergraduate programs of the country through a long exam, which is applied on a regular basis and has publicly available results, Brazil is a promising test scenario for bias issues related to overconfidence in students.

Our main results suggest that better performing students in Brazil "know what they do not know": performance is inversely correlated with overconfidence in both parts of the exam (general and specific knowledge). Senior students were also more able to assess their real performance.

Surprisingly, the female students presented a higher probability of being overconfident when compared to the male students. In addition, the students with higher incomes exhibited greater overconfidence, as well as those who spent less time to complete the exam.

The presented study opens space for new studies in the area of education, using a large database of students in a neutral environment.

References

- Armor, D.A. and Taylor, S.E. (2002). When predictions fail: The dilemma of unrealistic optimism. *Heuristics and biases: The psychology of intuitive judgment*, 334–347.
- Baker, L.A. and Emery, R.E. (1993). When every relationship is above average: Perceptions and expectations of divorce at the time of marriage. *Law and human behavior*, 17(4):439.

- Barber, B.M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics*, 116(1), 261–292.
- Beckmann, D., & Menkhoff, L. (2008). Will women be women? Analyzing the gender difference among financial experts. *Kyklos*, 61(3), 364–384.
- Bengtsson, C., Persson, M., and Willenhag, P. (2005). Gender and overconfidence. *Economics letters*, 86(2):199–203.
- Clayson, D.E. (2005). Performance overconfidence: metacognitive effects or misplaced student expectations? *Journal of Marketing Education*, 27(2):122–129.
- Dahlbom, L., Jakobsson, A., Jakobsson, N., and Kotsadam, A. (2011). Gender and overconfidence: are girls really overconfident? *Applied Economics Letters*, 18(4):325–327.
- Deaves, R., Lüders, E., & Luo, G.Y. (2008). An experimental test of the impact of overconfidence and gender on trading activity. *Review of finance*, 13(3), 555–575.
- INEP (2017). ENADE Microdata.
- INEP (2018). Preliminary Course Concept (CPC).
- Kennedy, E. J., Lawton, L., and Plumlee, E. L. (2002). Blissful ignorance: The problem of unrecognized incompetence and academic performance. *Journal of Marketing Education*, 24(3):243–252.
- Koku, P. S. and Qureshi, A. A. (2004). Overconfidence and the performance of business students on examinations. *Journal of Education for Business*, 79(4):217–224.
- Langendyk, V. (2006). Not knowing that they do not know: self-assessment accuracy of third-year medical students. *Medical education*, 40(2):173–179.
- Lichtenstein, S. and Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational behavior and human performance*, 20(2):159–183.
- Miller, T.M. and Geraci, L. (2011). Unskilled but aware: reinterpreting overconfidence in low-performing students. *Journal of experimental psychology: learning, memory, and cognition*, 37(2):502.
- Nowell, C. and Alston, R.M. (2007). I thought I got an A! overconfidence across the economics curriculum. *The Journal of Economic Education*, 38(2):131–142.
- Pallier, G. (2003). Gender differences in the self-assessment of accuracy on cognitive tasks. *Sex Roles*, 48(5–6):265–276.
- Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta psychologica*, 47(2):143–148.
- Taylor, S.E. and Brown, J.D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological bulletin*, 103(2):193.
- Weinstein, N.D. (1980). Unrealistic optimism about future life events. *Journal of personality and social psychology*, 39(5):806.