Goals and guesses as reference points: a field experiment on student performance

Gerardo Sabater-Grande[1] · Nikolaos Georgantzís[1,2] · Noemí Herranz-Zarzoso[3]

Abstract

In this paper, we study overconfidence and goal-setting in academic performance, with and without monetary incentives. Students enrolled in a microeconomics course were offered the possibility of setting their own target grade before taking part in the final exam. They were also asked to guess their grade immediately after they had taken the exam ("post-diction"). In general, students overestimated their performance, both at the goal-setting and at the post-diction stages. Controlling for several sources of this bias (cognitive abilities, academic record and self-reported academic confidence), we find that the use of monetary rewards mitigates the overestimation of potential achievements and eliminates overestimation of actual achievements through the improvement of actual performance. Our results suggest that monetary incentives do not cause subjects to put more effort into correct guesses but make them put more effort into academic performance. Using students' academic records to measure overall skill, we find a strong Dunning–Kruger bias which is intensified in the presence of monetary rewards.

Keywords Overconfidence bias · Reference points · Self-chosen goals · Post-dictions · Monetary incentives · Dunning · Kruger cognitive bias

1 Introduction

As has been established by prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), human behavior strongly depends on reference points which are used to assess whether an outcome is perceived as a gain or as a loss. Besides the

✉ Nikolaos Georgantzís
   nikolaos.georgantzis@bsb-education.com

1   LEE and Economics Department, Universitat Jaume I, Castellón, Spain

2   CEREN EA 7477, Burgundy School of Business, Université Bourgogne Franche-Comté, Dijon, France

3   Analysis Economic Department, University of Valencia, Valencia, Spain

🎄 Springer

reference points which may be dictated by social norms, individual reference points may relate to rational expectations (Koszegi & Rabin, 2006, 2007, 2009) and goals (Diecidue & Van De Ven, 2008). Using the properties of the prospect theory value function, Locke and Latham (2002) and Heath et al. (1999) suggest that goal-related reference points affect individual intrinsic motivation, improving actual performance.[1] Wu et al. (2008) present a model in which subjects' performance is improved by exogenously set goals. This finding was empirically supported in Allen et al. (2016) and Markle et al. (2018), with a sample of marathon runners. In the field of education, Meng (2019) tested the effect of grade aspiration-driven reference points on student performance.

In this paper, we study 2 types of reference points, depending on the timing of elicitation: self-chosen goals, elicited ex ante, by asking students to set their own target grade in a forthcoming exam and post-dictions, elicited immediately after the exam, by asking students to forecast their grade, given their perceived performance.[2] Following Fryer and Elliot (2008), self-chosen goals are empowered and proactive, creating commitment and acceptance. Thus, as personal bests, they act as reference points by inducing effort when current performance might otherwise be insufficient (Anderson & Green, 2018).[3] When goals are associated with monetary incentives, the performance improvement is even higher.[4]

Whereas goals can be considered as target-based reference points, post-dictions can be used as actual behavior-based reference points. This is so, given that "the expected value of an outcome is an easily available integrated mechanism that could be used as a (…) reference point" (Hack & von Bieberstein, 2015). Generally, the literature suggests that students' post-dictions of performance are more accurate than any type of prediction or target.[5] This is because, "whereas predictions are made prospectively and are based on what students think they know, post-dictions are made retrospectively and reflect the student's experience of the test" (Hacker et al., 2008).

In this vein, we are interested in analyzing the effectiveness of monetary incentives to encourage students to make a more thoughtful assessment of their potential and actual academic performance. We hypothesize that monetary incentives could reduce students' overestimation bias by improving their guesses. With the objective of testing this hypothesis, we conducted a randomized field experiment to elicit students' reference points, well before and immediately after the exam of a microeconomics course. In addition, we control for 2 potential factors driving students' reference points: their skill (potential or actual) and their self-reported academic confidence. Contrary to our hypothesis, we find that monetary incentives

---

[1] Their main properties are reference point dependence, loss aversion, and diminishing sensitivity.

[2] Experimental evidence suggests that goals and expectations are crucial in the formation of reference points (Hack et al., 2016).

[3] In words of van Lent and Souverijn (2020), "Economic theory (…) have shown how (…) goals can be used as reference points in order to increase performance for loss averse agents or hyperbolic discounters."

[4] See Goerg and Kube (2012), Dalton et al. (2015), Corgnet et al. (2015), Brookins et al. (2017) and Herranz-Zarzoso and Sabater-Grande (2018).

[5] See Glenberg and Epstein (1987), Maki (1998), Winnie and Jamieson-Noel (2002), and Serra and DeMarree (2016).

do not improve students' guesses but their academic performance, causing a significant decrease in students' overconfidence.

2 Literature review

This paper contributes to 3 lines of research: (1) monetary incentives and academic performance, (2) monetary incentives and overconfidence and, (3) overconfidence and cognitive ability.

First, we deal with the effect of monetary incentives on academic performance. The empirical evidence available[6] offers mixed conclusions on the potential of these interventions. Depending on the context, the students' age or implementation details, impacts have been generally modest or even null. As a departure point of the present study, in Herranz-Zarzoso and Sabater-Grande (2018), we conducted a randomized field experiment aiming at improving academic performance in a different course (Introductory microeconomics). In that study, we found that monetary incentives offered on the basis of self-chosen goals were effective to significantly increase students' grades.

Second, we focus on overconfidence bias. It has been generally observed that a person's subjective confidence in own performance is greater than the actual performance, resulting in the well-known overconfidence phenomenon.[7] Moore and Healy (2008) distinguish among 3 types of overconfidence: over-precision, a tendency of individuals to be excessively certain about the accuracy of their beliefs: overestimation, a tendency of subjects to exaggerate their absolute actual achievement or skill; and over-placement, which occurs when people perceive their performance in a group as better than it actually is.[8] Given the 2 types of reference point elicited, we observe 2 types of overestimation: students' overestimation of their targeted academic performance, defined as the difference between the grade obtained and the goal chosen before the exam, and students' overestimation of their actual academic performance, interpreted as the difference between the actual grade and the grade forecasted immediately after the exam.

Previous research has shown that students consistently overestimate their performance on academic exams,[9] especially when their grades are low. Specifically, Hacker et al. (2008) find that good students are usually more accurate, with a tendency to underestimate, while bad students usually overestimate their performance. Recently Santos-Pinto and de la Rosa (2020) reviewed empirical research on overconfidence and its effect on economic choices. Psychology evidence suggests that overconfidence depends on personal traits and environmental factors. Economic

---

[6] In the field of education, Lavecchia et al. (2016) offer a survey about the effectiveness of monetary incentives in improving academic performance.

[7] Overconfidence has been called "the most significant of the cognitive biases" by Kahneman (2011), Bazerman and Moore (2012) refer to it as "a mother of all biases" and Meloy et al. (2006) state that it "is one of the most robust findings in the decision and judgment literature".

[8] In Hoezl and Rustichini (2005) subjects estimate whether their skills are better than the median.

[9] See, for example, Chen (2003) and Foster et al. (2017), where it is concluded that "even after a great deal of practice (with over 10 exams), students on average did not improve the accuracy of their predictions".

evidence shows that overconfidence can persist even under monetary incentives and feedback.

Third, we deal with the relationship between subjects' cognitive ability and their overconfidence. Defining overconfidence as the difference between the guess and the proportion of correct answers,[10] a positive correlation between cognitive ability and confidence is found in Wolfe and Grosch (1990)[11] and Bruine de Bruin et al. (2007).[12] On the contrary, Stanovich and West (1998) reported a negative correlation of different cognitive ability tests with subjects' overconfidence. Later, Hoppe and Kusterer (2011) found that subjects with higher cognitive reflection test (CRT) scores had a significantly more precise self-assessment of their performance. However, the effects of cognitive abilities reported in earlier studies could depend on the type of overconfidence analyzed. In this vein, Duttle (2016) showed that, although overestimation[13] is not affected by cognitive abilities (as measured by a CRT), the CRT score was associated with a significant decrease in over-placement and over-precision. Nevertheless, Bialek and Domurat (2018) showed that the relationship between cognitive abilities and overconfidence disappeared after addressing 2 critiques: (1) the CRT does not measure cognitive abilities but, rather, the analytic cognitive style, and (2) overconfidence and cognitive ability are artificially correlated since the RPM test (which served as a basis for estimating overconfidence) is also a measure of cognitive abilities.

We are interested in testing for the so-called "Dunning–Kruger effect".[14] This well-known cognitive bias implies that when people are objectively unskilled in a given area, they tend to largely overestimate their knowledge. As Dunning (2011) states, this effect has been observed in multiple domains of skill[15] and knowledge, including academics. The empirical evidence available on the relation between skill and overconfidence is mixed, depending on the type of overconfidence analyzed and the methods used to measure the 2 variables.[16] Specifically, when the same task is used to measure overconfidence and skill, the empirical evidence obtained can be distorted by 'regression to the mean' effects, that is, individuals with higher skill are more likely to show less overconfidence. To avoid this effect, we use different tasks to measure overconfidence and skill, evaluated through a cognitive ability test or by means of the students' academic record.

---

[10] The problem with this overconfidence measure, as Moore and Healy (2008) suggest, is that being excessively sure of having answered correctly reflects both overestimation of your performance and excessive confidence in the precision of your knowledge.

[11] Using verbal and mathematical Scholastic aptitude test (SAT) scores to measure cognitive ability and a trivial quiz where subjects report their confidence in having answered correctly.

[12] Using Raven's standard progressive matrices for nonverbal cognitive ability and form G of the Nelson-Denny reading comprehension subtest for verbal ability.

[13] Defined as the difference between a subject's estimated number of correct test items in the Raven progressive matrices (RPM) test and his/her actual performance in the test.

[14] Kruger and Dunning (1999) find that the more poorly participants actually performed in tests of humor, grammar, and logic grossly, the more they overestimated their own performance.

[15] Park and Santos-Pinto (2010) find support for the Dunning-Kruger hypothesis in both poker and chess tournaments.

[16] See Feld et al. (2017).

In addition, we examine the relationship between perceived skill, measured by means of reported academic self-confidence, and overconfidence. Following Sander and Sanders (2009), we use the academic behavioral confidence scale[17] (ABC, hereafter) as a global measure of perceived academic confidence. Using a group test of general mental ability by Tandon (1971) and a self-confidence inventory by Agnihotri (1986), Dhall and Thukral (2009) investigated the relationship among intelligence, self-confidence, and academic achievement in schools in Pakistan. They found that intelligence was positively correlated with both self-confidence and academic achievement. However, Saenz et al. (2019) obtained that attendance, study habits/preparation, and/or prior performance did not offer a strong or robust explanation of students' grade predictions.

The rest of the paper is structured as follows: first, we introduce the design of the experiment implemented and the hypotheses; second, we present the methods used; third, we analyze the empirical evidence collected and present our results; last, we discuss them.

3 Experimental design

A randomized field experiment was conducted to analyze students' overconfidence using monetary incentives as the treatment variable. In the beginning of the semester, 154 students enrolled in a microeconomics course at a Spanish university in 2018.[18] They were offered the possibility of setting their reference points well before[19] and right after completion of the final exam. In the call, students were informed that those responding affirmatively would be immediately randomly assigned[20] to one of two groups: 1 without monetary incentives (NMI) and 1 with monetary incentives (MI). Additionally, we notified that participants would receive information about their corresponding group before they were invited to choose a goal for their final exam grade. From the 138 volunteers, 64 were randomly assigned to the NMI condition and 74 to the MI one. However, only 42 (16 females and 26 males) in the NMI treatment and 58 (26 females and 32 males) in the MI treatment decided to finally take the exam.[21] Furthermore, 16 students not responding affirmatively to our call were included in a baseline condition (non-participants group, NP group hereafter) to compare non-participants' final grades with those corresponding to the ones obtained by participating students. In doing so, we can check for potential self-selection bias. Table 1 summarizes the characteristics of the 3 groups.

---

[17] This scale encompasses cognitive judgments regarding study-related behaviors, such as confidence in independent study, attaining grades, discussing course material and questions, and academic engagement through attending formal sessions.

[18] The experiment was approved in the framework of an educational innovative project promoted by the university and approved by the faculty of law and economics.

[19] After a midterm exam (3 months before the final exam), self-chosen goals were elicited. Moreover, although we allowed their revision until 1 day before examination, no student submitted a new goal after their initial decisions. The midterm (final) exam represented 30% (70%) of the course grade.

[20] A pure random draw using a die was used to allocate subjects individually into groups.

[21] A Mann–Whitney test shows no gender differences between treatments (p value: 0.5029).

Table 1 Summary of groups

Groups	Number of subjects	Payment offered in the call	Willingness to participate	Randomization	Payment used
NP	16	No payment/money	No	No	–
NMI	42	No payment/money	Yes	Yes	No money
MI	58	No payment/money	Yes	Yes	Real money

In the call, students were instructed that, if assigned to the MI group, their monetary reward (R) would depend on the chosen reference point (RP) and the grade (GR) they obtained in the exam, according to the following scoring rule:

$$R = RP^2 \text{ if } GR = RP; R = 0 \text{ if } GR < RP$$

This is a more conservative test of overconfidence bias than a grade forecast question because the reward depends on a chosen reference point instead of the obtained grade. This is so because, like in Park and Santos-Pinto (2010),[22] a risk neutral player who overestimates her performance in a given amount incurs in a larger loss (0 earnings and loss equal to the square of the RP) than if she underestimates it in the same quantity (loss equal to the square of the difference between GR and RP). Thus, the optimal reference point of a risk neutral player should be smaller than his optimal grade forecast. To counterbalance the abrupt fall in earnings from a grade below the reference point, the quadratic scoring rule chosen here gives a good incentive for choosing a higher reference point because earnings increase marginally more for higher (successful) bets.

Moreover, students were instructed that only 1 of the 2 elicited reference points (participants' self-chosen goal or post-diction), chosen at random, would be used to determine their rewards. The actual average payment received from successful subjects in the experiment was €16 from a maximum potential reward of €49 (given that the top score in the final exam is 7 points). To distinguish between the 2 types of overconfidence, we define students' potential overconfidence (POC) as the self-chosen goal minus the grade obtained, and students' actual overconfidence (AOC) as the difference between the post-diction and the actual grade.

In addition, participants' cognitive ability was measured by means of the abstract reasoning part of the differential aptitude test (DAT-AR for PCA, Bennett et al., 1974). Moreover, we elicited scores on the ABC scale.[23] The research by Nicholson et al. (2013) suggested that undergraduates' confidence in their ability is related to academic performance. Specifically, this study showed that students who, at the beginning of the semester, were confident about their grades also perform better in

---

[22] They use a similar quadratic scoring rule as well as bets to incentivize poker and chess players' forecasts of relative performance in tournaments. The exam grade performance studied here resembles more the chess rather than the poker treatment, because students have a rather good idea of their abilities and would not be expected to treat their exam grades as purely random.

[23] See Sander and Sanders (2009).

their end-of-semester marks. Given the type of course considered in this study, we are especially interested in two factors of the ABC scale: grades and study beliefs.

Using the aforementioned experimental design, we propose four hypotheses. Because monetary incentives induce individuals' more thoughtful guesses, they should help to bring goals and expectations closer to actual academic performance. In consequence,

Hypothesis 1 (H1)

Introducing monetary incentives to elicit students' reference points, should soften both potential and actual overconfidence by improving guesses.

In addition, it is expected that monetary incentives will lead to more effort, producing a higher academic performance. Hence,

Hypothesis 2 (H2)

Implementation of monetary incentives will increase actual grades.

Moreover, assuming that both cognitive ability and academic record are good predictors of academic aspiration, performance and lower overconfidence, we hypothesize that:

Hypothesis 3 (H3)

Students with higher potential and actual skills will choose higher goals, they will achieve a better academic performance and they will show a lower level of overconfidence.

Finally, since self-reported studying and grade confidence should be related to students' aspirations,

Hypothesis 4 (H4)

Students self-reporting a higher academic confidence should choose higher goals.

Summing up, by controlling for potential driving factors, such as skill and reported academic self-confidence, our experimental design aims at analyzing whether both overconfidence and Dunning–Kruger bias can be mitigated using monetary incentives to elicit students' reference points.

4 Methods

In this section, we offer detailed information on the measures used to elicit students' cognitive abilities and self-reported academic confidence, respectively.

1. The abstract reasoning part of the differential aptitude test for personnel and career assessment. The abstract reasoning (AR) scale of the DAT used in this experiment is included in the Spanish adaptation of DAT-5 by the publisher TEA (Cordero & Corral, 2006). This test is used as a non-verbal measure of reasoning ability and involves the capacity to think logically and to perceive relationships in figures made up of abstract patterns. It is considered as a marker of fluid intelligence (Colom et al., 2007) and the component of intelligence most related to general intelligence or g factor (McGrew, 2009). The advantage of this test is that it can be administered quickly, containing 40 multiple-choice items within a 20-min time limit.

2. The academic behavioral confidence scale (Sander & Sanders, 2009). The ABC scale used in this research was the 24-statement version. These statements elicit

the student's expectation of achieving good grades in assessments (grades), engaging in independent study (studying), attending lectures, tutorials and other taught sessions (attendance), and discussing material with tutors, lecturers, and peers (verbalizing). However, all analyses presented in this paper were computed only for 2 ABC subscales: grades and studying. In Sander et al. (2011), the ABC scale shows cross-cultural validity when translated into Spanish and administered to over 2 thousand Spanish psychology students.

5 Data analysis

5.1 Sample self-selection

Given that our design requires students' willingness to participate in the experiment, potential self-selection problems do not affect differences between MI and NMI. Nevertheless, the baseline group allows us to test for self-selection bias by comparing prior midterm grades of participating and non-participating subjects. A Mann–Whitney test shows that differences between participant and non-participant groups are not statistically significant (p value: 0.1517).

5.2 Descriptive statistics and tests

Table 2 shows the sample split into NMI and MI, and presents descriptive statistics of: normalized data to a scale from 0 to 10 corresponding to (1) subjects' self-chosen goals, (2) post-dictions, (3) grades and (4) POC and AOC. Moreover, we display additional descriptive statistics corresponding to: (1) score in the DAT-AR test, (2)

Table 2 Descriptive statistics and group comparisons of self-chosen goals, post-dictions, grades, overconfidence, DAT-AR score, studying and grades subscales of the ABC scale, and academic record

	N NMI	N MI	Mean NMI	Mean MI	S.D NMI	S.D I	M–W test/ T test p value	n ex-post power
Self-chosen goal	42	58	6.01	5.83	0.98	0.99	0.4348	942
Post-diction	42	58	5.34	5.32	1.58	1.23	0.7625	157,344
Grade	42	58	4.04	5.26	2.23	2.41	0.0116**	
POC	42	58	1.97	0.58	2.16	2.40	0.0037***	
AOC	42	58	1.30	0.06	2.06	2.32	0.0072***	
DAT-AR score	42	58	25.19	22.74	7.62	7.71	0.1186	308
Studying confidence	42	58	14.31	15.10	2.87	2.16	0.1539	326
Grades confidence	42	58	21.62	22.03	3.81	4.02	0.5320	2866
Academic record	42	58	6.40	6.40	0.66	0.49	0.4402	–

***$p<0.01$, **$p<0.05$, *$p<0.1$

scores in the grades and studying subscales of the ABC scale, and (3) academic record. Additionally, we present p values from a t test or a Mann–Whitney (M–W) test comparing[24] NMI and MI conditions for all these variables. Moreover, to check whether the lack of significance for some differences was due to a low statistical power, we conduct an ex-post power analysis using Stata with power set at 0.80 and probability at 0.05. The last column of Table 2 shows the minimum sample size to find statistically significant differences.

We observe that although self-chosen goals and post-dictions are similar in the NMI and the MI groups, both types of overconfidence are significantly higher when subjects' reference points are elicited without incentives (NMI group). Specifically, a POC of almost 2 points is reduced to a quarter when subjects' self-chosen goals are elicited using money. In the same manner, an AOC of 1.3 points disappears when subjects' post-dictions are obtained under monetary incentives. It is worth mentioning that this reduction of the students' overconfidence is not caused by a difference in their expectations but, rather, by an improvement in their performance. Moreover, this reduction in both POC and AOC is not influenced by cognitive ability or academic confidence/records, since these variables do not present significant differences across groups.

In Fig. 1, means are presented and statistical differences are tested through a Wilcoxon test. Specifically, in the first row, we display means of self-chosen goals, post-dictions, and grades, split between NMI and MI.

For both groups, we obtain that the median of students' self-chosen goals is significantly higher than the median of their post-dictions. However, only NMI subjects obtain a median grade significantly lower than their median post-diction, since MI subjects post-diction their grade accurately. The 2nd row of Fig. 1 presents the mean values of POC and AOC for both groups of subjects, showing that real monetary incentives significantly reduce both POC and AOC medians, the latter being close to 0.

Figure 2 displays the confidence measured using self-chosen goals and post-dictions against the grade obtained in the midterm exams. The 45 ° line provides a benchmark, given that points above the line would represent overconfidence whereas points below the line would represent under-confidence. The dashed line is the minimum grade required to pass an exam. In the right panel of this figure, we can observe that most of the NMI subjects show both POC and AOC. Thus, this right panel reproduces the effect of overconfidence in general terms. This pattern is not observed for MI subjects (left panel). In fact, they show overconfidence for low grades and under-confidence for high grades, as can be expected when a regression to the mean effect is present.

Now, we analyze the relationship between subjects' confidence and cognitive ability. Figure 3 displays both potential and actual confidence for subjects included in both the NMI and the MI group against cognitive ability. For both groups, NMI and MI subjects, Fig. 3 shows no pattern relating confidence and cognitive ability.[25]

---

[24] We use t tests for comparisons between normally distributed variables (grade, POC and AOC, and DAT-AR score) and the M–W test for comparisons between the remaining (non-normal) variables.

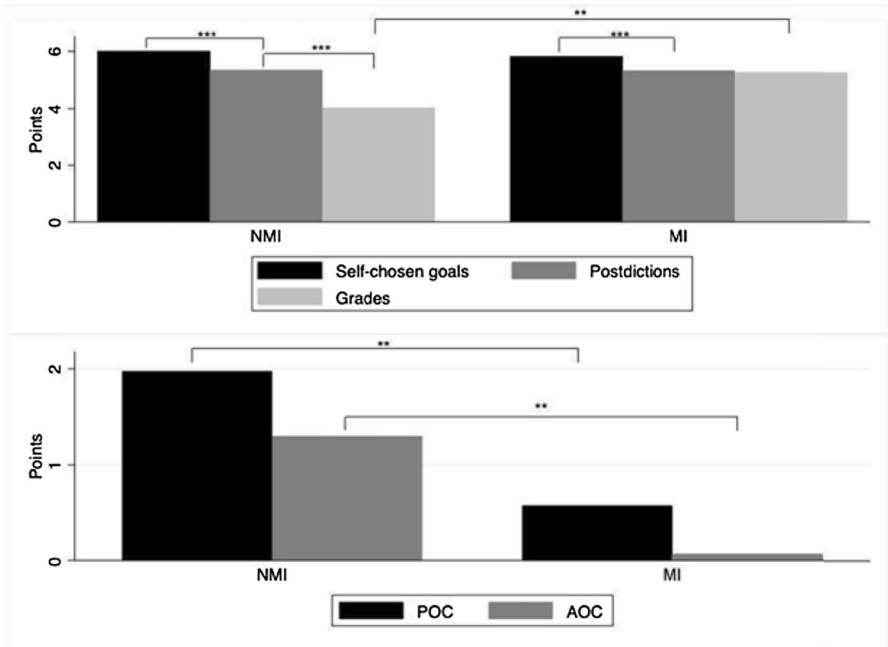[25] Spearman's coefficients show a non-significant correlation.

Fig. 1 Self-chosen goals, post-dictions, grades, POC, and AOC means for the final exam. ***$p<0.01$, **$p<0.05$, *$p<0.1$
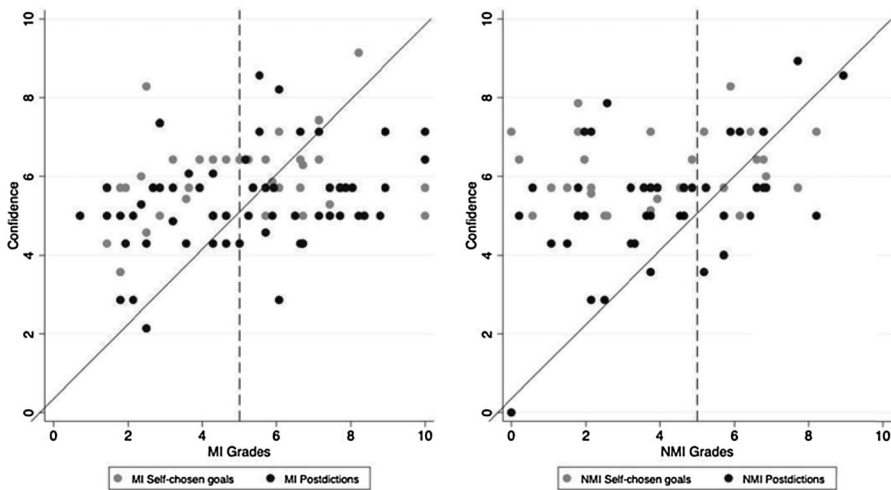

Fig. 2 Confidence using self-chosen goals and post-dictions against grades in the final exam

Moreover, for both MI and NMI samples, subjects were divided into 2 groups according to their reasoning ability. A subject was classified as "high (low) reasoning" if her score was higher (lower) than the median score in the DAT-AR test.
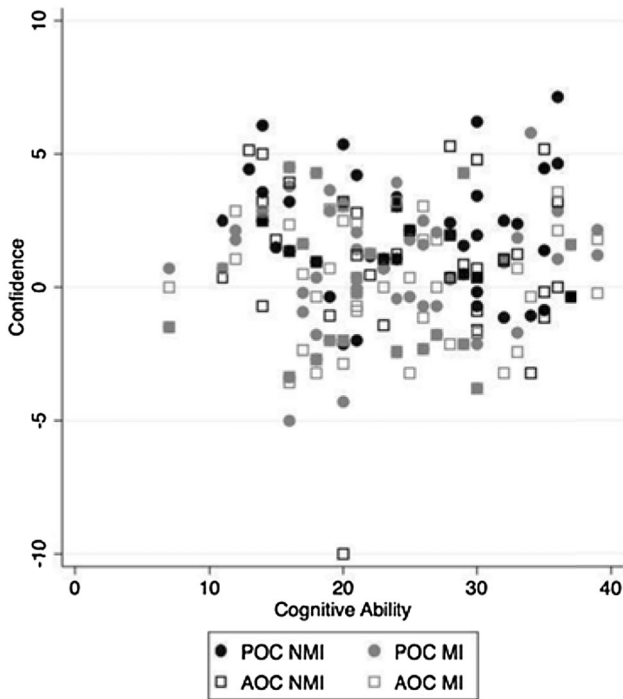
Fig. 3 Confidence using self-chosen goals and post-dictions against grades in the final exam

Using a Mann–Whitney test[26] we obtain that high reasoning incentivized students choose higher goals than low reasoning ones. However, there are not significant differences for post-dictions, grades and overconfidence between high reasoning and low reasoning participants for both samples, incentivized and non-incentivized students.

The same procedure was implemented using academic record, and studying and grades confidence variables. Regarding academic record, we find that incentivized students with better academic records do not set more ambitious goals or higher post-dictions but obtain better grades resulting in a significant decrease in both potential and actual overconfidence.[27] This decrease is not significant[28] for the non-incentivized sample.

---

[26] At the 10% level.

[27] Mann–Whitney test p values obtained comparing incentivized high academic record students and low academic record ones corresponding to goals, post-dictions, grades, potential and actual overconfidence are 0.2756, 0.9144, 0.0101, 0.0242 and 0.0192 respectively.

[28] Mann–Whitney test p values, comparing potential and actual overconfidence between non-incentivized high academic record students and low academic record ones, are 0.1170 and 0.1682 respectively.

Finally, in relation to self-reported confidence, we obtained no significant[29] differences between reference points, grades and overconfidence chosen by incentivized subjects with a higher studying/grade confidence and those with a lower one. However, non-incentivized highly self-confident students obtain significantly[30] better grades than those with a lower self-confidence, resulting in a significantly[31] lower overconfidence.

5.3 Regression analysis

Reference points

In this section, potential explanatory factors of self-chosen goals and post-dictions are explored. OLS models are estimated to explain both reference points elicited.[32] The potential driving factors used are: (1) cognitive ability, (2) academic record, (3) self-reported grades confidence and studying confidence, and (4) gender (1, if the student is a woman and 0, student is a man).

The first important result is related to cognitive ability. The direct relationship between cognitive ability and self-chosen goals only occurs when participants are monetarily rewarded. Moreover, cognitive ability does not play a role in explaining post-dictions independently of the type of incentives offered. In addition, subjects with a better academic record only report higher self-chosen goals and post-dictions when they are not monetarily incentivized. In addition, a higher reported academic confidence does not lead to higher goals or post-dictions. Regarding the importance of the type of incentives used to elicit students' reference points, we find that the MI dummy is not significant: self-chosen goals and post-dictions by MI subjects are, on average, similar to those of NMI subjects. Finally, the reference points elicited do not differ across genders.

Therefore, contrary to H1, we can state that:

Result 1: Reference points are not affected by monetary incentives.

In addition, H3 is only partially confirmed in relation to subjects' skills and H4 is rejected:

Ancillary result 1: Under monetary incentives, we find a direct relationship between cognitive ability and subjects' ambition choosing their goals. However, reference points are not affected by students' academic record and reported academic confidence.

---

[29] Mann–Whitney test p values, comparing incentivized high studying (grades) confidence subjects and low confidence subjects regarding to goals, post-dictions, grades, potential and actual overconfidence, are 0.8705 (0.5051), 0.9385 (0.9730), 0.5562 (0.1767), 0.6438 (0.2102) and 0.6882 (0.2358) respectively.

[30] Mann–Whitney test p value comparing grades between non-incentivized high studying (grades) confidence subjects and low confidence subjects is 0.0101 (0.0175).

[31] Mann–Whitney test p values, comparing potential and actual overconfidence between non-incentivized high studying (grades) confidence subjects and low confidence subjects, are 0.0175 (0.1084) and 0.0903 (0.0747) respectively.

[32] Given that our dependent variables are limited and as Park and Santos-Pinto (2010) point out, this fact produces heteroskedasticity, invalidating the statistical inference. Thus, we use robust standard errors in the OLS regressions. Additionally and following Park and Santos-Pinto (2010), to check whether our results are robust, we perform a logit transformation of the dependent variables to show the robustness of our results (see Appendix A2).

5.3.1 Grades

In this section, determinants of students' grades are analyzed. Below we present OLS models explaining students' grades.[33]

The first model shows that the student's academic record plays a crucial role in the grades obtained. The last 2 models indicate that the predictive power depends crucially on monetary incentives. In addition, we obtain that subjects' confidence in their study method positively affects grades, especially in the absence of monetary incentives. Moreover, we obtain that a higher cognitive ability predicts a better academic performance only for the NMI group. Regarding the size of the reference points elicited, only students' post-dictions were related with their grades, but the predictive power disappeared when they were monetarily incentivized. MI subjects obtain, on average, 1.08 points more than NMI subjects. Thus, real monetary incentives are effective as a means of significantly improving students' academic performance.

Thus, confirming H2, we can state that:

Result 2: Students' academic performance is improved by monetary incentives.

Moreover, confirming partially H3 regarding actual skills, we find that:

Ancillary result 2: Especially under monetary incentives, we find a direct relationship between academic record and grades.

5.3.2 Overconfidence

In this section, 4 OLS models are estimated to shed light on the determinants of subjects' overconfidence.

Regressions indicate that monetary incentives significantly reduce both POC and AOC. Specifically, in the presence of monetary incentives POC and AOC are on average 1.12 points lower than in the absence of monetary incentives. In conclusion:

Result 3: By improving grades, monetary incentives reduce overconfidence in goal-setting and make it disappear in post-diction of grades.

In addition, from the OLS models, we obtain that students' academic records have a negative effect on both POC and AOC. That is, subjects with a better academic record show less overconfidence, especially when they are incentivized with money. The same pattern is found regarding the studying confidence scale, but only in the absence of monetary incentives. Therefore, subjects with more confidence in their study methods show lower overconfidence. In contrast, cognitive ability and gender do not explain POC or AOC.[34]

Concerning the existence of a Dunning–Kruger bias, we can summarize our findings as follows. Our results confirm this phenomenon using cognitive ability as the measure of potential skill only when no incentives were offered. However, this effect disappears when monetary incentives are used. Specifically, although students

---

[33] Because the dependent variable has a limited variation, robust standard errors are used on OLS regressions. Furthermore, the results are robust to the transformation of the dependent variable (see appendix A2).

[34] With one exception: a significant negative relation between cognitive abilities and AOC is found when subjects are not monetarily incentivized.

with higher cognitive abilities choose higher goals, we do not find any relationship between cognitive ability and potential overconfidence, when self-chosen goals are elicited using monetary incentives. In addition, fluid intelligence of MI subjects is not related to their post-dictions or to their actual overconfidence. Thus, we can state that, under monetary incentives and using cognitive ability as measure of subjects' (potential) skill, we do not find any evidence of the Dunning–Kruger effect.

However, when we use students' academic record to measure their (actual) skill, we obtain that monetary incentives reinforce the presence of the Dunning–Kruger bias,[35] observed to a lesser extent in the absence of money rewards. Therefore, we partly confirm H3 regarding overconfidence:

Ancillary result 3: Using monetary incentives to elicit students' reference points, we (do not) find an inverse relationship between actual (potential) skills and overconfidence.

Lastly, we reject H4:

Ancillary result 4: When reference points are elicited without monetary incentives, although there is no direct relationship between studying confidence and goals, there is a direct (inverse) one between studying confidence and grades (overconfidence). Moreover, the self-reported confidence in their grades does not correlate with their reference points, grades or overconfidence in any case.

6 Discussion

To the best of our knowledge, this paper is the first to study students' overconfidence using monetary incentives to analyze self-chosen goals and post-dictions as reference points. We consider money as an effective incentive because, it is a non-satiable good and, as Croson (2005) points out, "everyone values it, in contrast with extra-credit points or other grade-related rewards which may be valued only by students who are grade-conscious and/or whose grade may be affected by the outcome" and it is a non-satiable good.[36]

Both students' self-chosen goals and post-dictions are often elicited in the literature with non-incentive compatible methods. Following Murstein (1965), multiple survey data confirming overconfidence have been collected using no incentives.[37] However, to motivate students in their task of forecasting, in the past decade, some contributions have introduced 2 types of incentives: extra grade points and money. Miller and Geraci (2011), Magnus and Peresetsky (2018) and Caplan et al. (2018) use bonus points to encourage students to reveal their honest guesses

---

[35] This result is in line with Park and Santos-Pinto (2010) where players' ratings were used as a measure of their skill.

[36] Effectiveness of the joint action of self-chosen goals and real monetary incentives in higher education is shown in Herranz-Zarzoso and Sabater-Grande (2018).

[37] Examples are Kruger and Dunning (1999), Grimes (2002), Falk and Knell (2004), Svanum and Bigatti (2006), Andrews et al. (2007), Nowell and Alston (2007), Burns (2007), Jensen and Moore (2008), Khachikian et al. (2011), Armstrong (2013), Hossain and Tsigaris (2013, 2015), Sackett et al. (2014)Query, Serra and DeMarree (2016), Sturges et al. (2016), Foster et al. (2017) and Clark et al. (2020).

about grades. In general, this type of incentive fails to show improvement in students' expectations or performance.

Monetary incentives were introduced in Feld et al. (2017) and Saenz et al. (2019).[38] Ehrlinger et al. (2008) and Gutiérrez and Schraw (2015) are among the few papers analyzing the role of monetary incentives in students' predictions with mixed results. Specifically, Ehrlinger et al. (2008) find that even offering $100 to college students who are exactly correct in their prediction did not lead to more accurate estimates of the number of questions answered correctly. Gutiérrez and Schraw (2015) used a monetary reward of US$10, contingent upon meeting or exceeding the test performance criterion at posttest. They find that incentives improve calibration accuracy only when combined with a training strategy.

Our results contrast with this previous literature suggesting the importance of monetary incentives as a means of reducing students' overconfidence through the improvement of performance. We find that monetary incentives do not cause students to put more effort into correct guesses but, rather, in exam performance. In particular, our results suggest that setting goals is more motivating in the presence of monetary incentives, making goals and actual performance converge by enhancing the latter.

Specifically, taking into account individual characteristics like skill and reported academic confidence, we obtain that students' overestimation of their potential achievements is significantly reduced when money is used to elicit their self-chosen goals. Rather than revising their goals down to match a lower skill, subjects brought their performance up to meet their aspirations. This effect is stronger in the case of subjects' actual overestimation of their grades, eliminating the bias.

Our study presents some strengths and undoubtedly several limitations. Among the major strengths: (1) we use non-negligible monetary incentives to elicit students' self-chosen goals and post-dictions; (2) we control for potential driving factors like cognitive ability, academic record and self-reported academic confidence; and (3) we double-check potential self-selection effects in our sample. Our study also had some limitations including: (1) limited sample size; (2) incentive effects dependent on the quadratic reward function introduced in our experimental design; and (3) uncontrolled factors which may partially be responsible for some of the differences reported here. With reference to the first limitation, using an ex-post power analysis, we check whether the lack of significance of some differences between treatment groups was due to a small sample size. The results show that samples in each treatment group would have to increase to at least 942 subjects to find statistically significant differences at a 5% level in self-chosen goals and to 157,344 in post-dictions. In relation to the 2nd reservation, the scoring rule used here is a conservative test for overconfidence. So, mitigation effects of monetary incentives could be smaller if subjects were rewarded by alternative scoring rules. Regarding the last concern, other uncontrolled factors may partially be accountable for some of the effects reported here. Thus, our results have to be interpreted cautiously when

---

[38] The former study adopted a lottery draw in which students could win one of two gift vouchers worth €20 if their prediction four weeks before the exam was within 0.25 points of their grade. However, the latter paper uses a low $50 prize for one in about 40 participants.

establishing causal relationships between monetary incentives and academic performance.

Although this paper reveals the positive effect of monetary incentives in academic performance, policymakers might be concerned by (1) crowding-out effects (Gneezy et al., 2011) and (2) the financial resources needed. Regarding the argument that monetary incentives could crowd out students' intrinsic motivations, List et al. (2018) only found a limited temporary effect, obtaining that one year later non-incentivized tests were not negatively affected. With respect to the second concern, Herranz-Zarzoso and Sabater-Grande (2018) showed that lower cost-intensive mechanisms as a rank-order tournament scheme can be as effective as a piece-rate payment mechanism like the one adopted here.

Further research is needed to explore factors like personality traits to better explain subjects' overconfidence bias, and new tools reinforcing monetary incentives to soften this overconfidence, like experience and feedback on previous academic tasks.

7 Appendix

7.1 Experimental instructions (translated from Spanish)

7.1.1 Call instructions

The LEE Research Team of the University Jaume I is conducting a study to evaluate which factors contribute to good performance on this course. The team will use your responses, grades from this course and your academic records if you consent to participate. All the information will be anonymously associated with the findings of this research. You have the right to withdraw from the study at any time during the semester. If you withdraw there will be no consequences for you; your academic standing and record will not be affected.

If you consent to participate in this study you will be immediately randomly assigned to one of two groups: group 1 or group 2. If you have been selected in group 1 you will be paid by your decisions but in the case of being assigned to group 2 you will not be rewarded by them. You will receive information coming soon by email about your corresponding group, before you have to make any decision. Once you have been informed, you will have to set a goal for your final exam grade in this course. Remember that the final exam is worth 70% of your final course grade (7 points is the maximum grade to be achieved). You will be allowed to revise your goal until one day before examination. All communications should occur over email.

Moreover, you will be asked to guess the grade you think you will get in the final exam immediately after being completed. Only one of the two decisions, chosen at random, will be used to determine your reward if you have been randomly selected in group 1. In this case, your reward, in euros will be equal to:

$$R = \text{Guess}^2 \text{ if Grade} \geq \text{Guess}$$

In addition, if you have declared your willingness to participate, you will be asked to perform two additional tasks (a 20-min abstract reasoning test and and a short question test).

Please respond to this mail if you consent to participate in this study.

7.1.2 Instructions for group 1 participants

You have been randomly selected in group 1. Thus, your decisions can be monetary rewarded. The first decision you need to make is choosing a goal for your final course grade (remember that your communication must be by email and the goal can be revised until one day before examination). Please, think carefully before setting your goal.

Remember that the second decision you must take is to guess the grade you think you will get in the final exam, immediately after being completed (you will be asked to write your guess at the final page of the exam). Only one of the two decisions, chosen at random, will be used to determine your reward. For transparency, we will use one number of the Lotería Nacional raffle of May 26, 2018 for each class group (A, B, C, D, E and F).

– Group A: third number of the first prize
– Group B: forth number of the first prize
– Group C: fifth number of the first prize
– Group D: first number of the second prize
– Group E: second number of the second prize
– Group F: first number of the second prize

For each course group, we will use your first decision to reward you in case that the raffle corresponding number is 1, 2, 3 or 4, and we will adopt your second decision to remunerate you if the raffle number is 5, 6, 7, 8 or 9.

Remember that, in both decisions, you have to guess your grade over 7 and if you equal or exceed your guess, you will received a monetary payoff in euros according to the following function:

$$R = \text{Guess}^2 \text{ if Grade} \geq \text{Guess}$$

7.1.3 Instructions for group 2 participants

You have been randomly selected in group 2. Thus, your decisions will not be monetary rewarded. The first decision you need to make is choosing a goal for your

final course grade (remember that your goal can be revised until one day before examination). The second decision you must take is to guess the grade you think you will get in the final exam immediately after being completed.

8 Logit transformation

Since reference points and grades are limited variables taking normalized values between 0 and 10, the results obtained by OLS can suffer from heteroskedasticity which invalidates statistical inference. To avoid this shortcoming, on the one hand, the standard errors reported in OLS regressions are robust to the presence of heteroskedasticity. On the other hand, to check if the results obtained are robust, we perform a logit transformation of our dependent variables based on reference points and grades. The logit transformation of self-chosen goals (SCG), post-dictions (POST), grades (G), potential overconfidence (POC) and actual overconfidence (AOC) are given by:

$$SCG_i = \ln\left(\frac{SCG_i}{10 - SCG_i}\right)$$

Table 3 OLS regressions for self-chosen goals and post-dictions elicited in the final exam splitting the sample between monetarily incentivized and non-monetarily incentivized subjects

	Self-chosen goals all subjects	Self-chosen goals MI subjects	Self-chosen goals NMI subjects	Post-dictions all subjects	Post-dictions MI subjects	Post-dictions NMI subjects
MI	− 0.0782 (0.190)			− 0.0681 (0.269)		
Cognitive ability	0.0314** (0.0144)	0.0384** (0.0183)	0.0141 (0.0234)	− 0.0105 (0.0191)	0.00393 (0.0223)	− 0.0425 (0.0364)
Academic record	0.477*** (0.168)	0.217 (0.255)	0.732*** (0.195)	0.676** (0.271)	0.549 (0.402)	0.801** (0.386)
Grades confidence	0.0306 (0.0278)	0.0235 (0.0308)	0.0411 (0.0537)	− 0.00464 (0.0359)	− 0.0325 (0.0372)	0.0642 (0.0682)
Studying confidence	− 0.0297 (0.0460)	0.0340 (0.0782)	− 0.0886 (0.0563)	0.0669 (0.0741)	0.123 (0.0890)	− 0.0138 (0.0947)
Gender	− 0.194 (0.205)	− 0.299 (0.275)	− 0.0843 (0.307)	− 0.369 (0.279)	− 0.347 (0.334)	− 0.526 (0.560)
Constant	2.088 (1.333)	2.675 (1.928)	1.382 (1.603)	0.631 (1.896)	0.739 (2.340)	0.299 (2.618)
N	100	58	42	100	58	42
R-squared	0.233	0.184	0.380	0.112	0.099	0.167

Table 4 OLS regression for students' grades

	Grades all subjects	Grades MI subjects	Grades NMI subjects
MI	1.079*** (0.376)		
Cognitive ability	0.0198 (0.0258)	– 0.0103 (0.0330)	0.0842** (0.0361)
Academic record	1.673*** (0.404)	2.179*** (0.555)	0.988* (0.492)
Grades confidence	– 0.0244 (0.0565)	– 0.0238 (0.0631)	– 0.126 (0.0931)
Studying confidence	0.295*** (0.0977)	0.268* (0.147)	0.401*** (0.144)
Gender	– 0.264 (0.423)	– 0.942 (0.603)	0.976** (0.479)
Self-chosen goal	0.0661 (0.247)	0.0672 (0.367)	0.322 (0.316)
Post-diction	0.359*** (0.126)	0.292 (0.247)	0.519*** (0.135)
Constant	– 14.15*** (2.527)	– 13.50*** (3.721)	– 12.50*** (3.330)
Observations	100	58	42
R-squared	0.439	0.386	0.581

The regression is split into monetarily incentivized and non-monetarily incentivized subjects. Robust standard errors in parentheses

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 5 OLS regressions for POC and AOC splitting the sample into monetarily incentivized and non-monetarily incentivized subjects

	POC	POC MI subjects	POC NMI subjects	AOC	AOC MI subjects	AOC NMI subjects
MI	– 1.122*** (0.399)			– 1.116*** (0.406)		
Cognitive ability	0.0127 (0.0287)	0.0448 (0.0353)	– 0.0527 (0.0392)	– 0.0286 (0.0293)	0.0115 (0.0360)	– 0.109*** (0.0372)
Academic record	– 1.471*** (0.436)	– 2.138*** (0.595)	– 0.908 (0.578)	– 1.272*** (0.426)	– 1.805*** (0.642)	– 0.839* (0.477)
Grades confidence	0.0558 (0.0641)	0.0557 (0.0698)	0.121 (0.109)	0.0196 (0.0667)	– 0.00217 (0.0712)	0.144 (0.0861)
Studying confidence	– 0.348*** (0.107)	– 0.273* (0.159)	– 0.454*** (0.137)	– 0.250** (0.114)	– 0.180 (0.161)	– 0.379** (0.143)
Gender	0.220 (0.463)	0.767 (0.680)	– 0.762 (0.569)	0.0402 (0.456)	0.710 (0.648)	– 1.206** (0.545)
Constant	15.88*** (2.934)	15.79*** (3.822)	13.29*** (3.847)	14.41*** (2.812)	13.81*** (4.220)	12.20*** (3.193)
Observations	99	57	42	99	57	42
R-squared	0.303	0.269	0.360	0.253	0.218	0.378

Robust Standard errors in parentheses ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 6 OLS regressions for self-chosen goals and post-dictions elicited in the final exam splitting the sample between monetarily incentivized and non-monetarily incentivized subjects

	Self-chosen goals all subjects	Self-chosen goals MI subjects	Self-chosen goals NMI subjects	Post-dictions all subjects	Post-dictions MI subjects	Post-dictions NMI subjects
MI	− 0.0262			− 0.0721		
	(0.0906)			(0.118)		
Cognitive Ability	0.0147**	0.0182**	0.00607	1.89e− 05	0.00205	− 0.00544
	(0.00665)	(0.00864)	(0.0105)	(0.00757)	(0.00986)	(0.0121)
Academic Record	0.231***	0.0857	0.372***	0.302**	0.249	0.350*
	(0.0833)	(0.111)	(0.0964)	(0.124)	(0.180)	(0.174)
Grades Confidence	0.0117	0.0104	0.0143	− 0.00408	− 0.0127	0.00962
	(0.0121)	(0.0134)	(0.0237)	(0.0152)	(0.0161)	(0.0326)
Studying Confidence	− 0.0141	0.00929	− 0.0366	0.0110	0.0526	− 0.0275
	(0.0214)	(0.0379)	(0.0252)	(0.0274)	(0.0390)	(0.0352)
Gender	− 0.0882	− 0.137	− 0.0356	− 0.146	− 0.156	− 0.165
	(0.0904)	(0.120)	(0.133)	(0.128)	(0.144)	(0.271)
Constant	− 1.404**	− 0.908	− 1.869**	− 1.658**	− 1.940*	− 1.633
	(0.625)	(0.847)	(0.757)	(0.743)	(1.045)	(0.982)
N	100	58	42	99	58	41
R-squared	0.233	0.171	0.417	0.122	0.103	0.179

Logit transformation implemented in the dependent variables

Standard errors in parentheses ***$p<0.01$, **$p<0.05$, *$p<0.1$

Table 7 OLS regression for students' grades. The regression is split into monetarily incentivized and non-monetarily incentivized subjects

	Grades all subjects	Grades MI subjects	Grades NMI subjects
MI	0.503*** (0.188)		
Cognitive ability	0.0115 (0.0130)	− 0.00439 (0.0162)	0.0418** (0.0197)
Academic record	0.792*** (0.210)	0.925*** (0.284)	0.642* (0.322)
Grades confidence	− 0.0182 (0.0286)	− 0.00708 (0.0291)	− 0.0854* (0.0500)
Studying confidence	0.140*** (0.0488)	0.112 (0.0734)	0.209** (0.0770)
Gender	− 0.0155 (0.207)	− 0.367 (0.296)	0.626** (0.267)
Self-chosen goal	0.0838 (0.122)	0.106 (0.176)	0.125 (0.184)
Post-diction	0.138* (0.0708)	0.103 (0.123)	0.234*** (0.0764)
Constant	− 9.197*** (1.409)	− 8.267*** (1.908)	− 9.049*** (1.998)
Observations	97	56	41
R-squared	0.379	0.316	0.521

Standard errors in parentheses ***$p<0.01$, **$p<0.05$, *$p<0.1$. Logit transformation implemented in the dependent variables

Table 8 OLS regressions for POC and AOC splitting the sample into monetarily incentivized and non-monetarily incentivized subjects

	POC	POC MI subjects	POC NMI subjects	AOC	AOC MI subjects	AOC NMI subjects
MI	− 0.247*** (0.0856)			− 0.237*** (0.0843)		
Cognitive ability	0.00361 (0.00619)	0.0103 (0.00737)	− 0.0101 (0.00923)	− 0.00576 (0.00600)	− (0.00728)	− 0.0230*** (0.00793)
Academic record	− 0.321*** (0.0925)	− 0.453*** (0.125)	− 0.209 (0.126)	− 0.270*** (0.0888)	− 0.377*** (0.134)	− 0.181* (0.101)
Grades confidence	0.0113 (0.0134)	0.0110 (0.0140)	0.0249 (0.0243)	0.00423 (0.0137)	− 0.000607 (0.0143)	0.0310 (0.0184)
Studying confidence	− 0.0756*** (0.0232)	− 0.0539 (0.0333)	− 0.102*** (0.0301)	− 0.0531** (0.0237)	− 0.0364 (0.0330)	− 0.0826** (0.0307)
Gender	0.0350 (0.0980)	0.157 (0.142)	− 0.178 (0.122)	0.00201 (0.0952)	0.144 (0.134)	− 0.259** (0.116)
Constant	3.460*** (0.641)	3.283*** (0.798)	3.009*** (0.864)	3.052*** (0.592)	2.865*** (0.874)	2.627*** (0.684)
Observations	99	57	42	99	57	42
R-squared	0.309	0.267	0.367	0.257	0.218	0.382

Robust standard errors in parentheses ***$p<0.01$, **$p<0.05$, *$p<0.1$. Logit transformation implemented in the dependent variables

$$POST_i = \ln\left(\frac{POST_i}{10 - POST_i}\right)$$

$$G_i = \ln\left(\frac{G_i}{10 - G_i}\right)$$

$$POC_i = \ln\left(\frac{POC_i}{10 - POC_i}\right)$$

$$AOC_i = \ln\left(\frac{AOC_i}{10 - AOC_i}\right)$$

Once the dependent variables have been transformed, we performed OLS regressions. Tables 3, 4 and 5 present the results in a similar way that Tables 6 7 and 8, showing that previous results are robust.

Availability of data and material Data are available and provided as supplementary information.

Code availability Not applicable.

Declarations

References

Agnihotri, R. (1986). *Manual for Agnihotri's self confidence inventory (ASCI)*. Berlin: National Psychological Corporation Ed.

Allen, E. J., Dechow, P. M., Pope, D. G., & Wu, G. (2016). Reference-dependent preferences: Evidence from marathon runners. *Management Science, 63*(6), 1657–1672. https://doi.org/10.1287/mnsc.2015.2417

Anderson, A., & Green, E. A. (2018). Personal bests as reference points. *Proceedings of the National Academy of Sciences of the United States of America, 115*(8), 1772–1776. https://doi.org/10.1073/pnas.1706530115

Andrews, K., Swanson, J., & Kugler, P. (2007). Grade Expectations. *Journal of Economoics & Economic Education Research, 8*(2), 3–18.

Armstrong, M. J. (2013). A preliminary study of grade forecasting by students. *Decision Science Journal of Innovative Education, 11*(2), 193–210. https://doi.org/10.1111/dsji.12003

Bazerman, M. H., & Moore, D. A. (2012). *Judgment in managerial decision making*. John Wiley & Sons.

Benjamin, D. J., Brown, S. A., & Shapiro, J. M. (2013). Who is "behavioral"? cognitive ability and anomalous preferences. *Journal of the European Economic Association, 11*(6), 1231–1255. https://doi.org/10.1111/jeea.12055

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1974). *Manual for the differential aptitude tests*. Psychological Corporation Ed.

Białek, M., & Domurat, A. (2018). Cognitive abilities, analytic cognitive style and overconfidence: A commentary on Duttle (2016). *Bulletin of Economic Research, 70*(1), 119–125. https://doi.org/10.1111/boer.12117

Brookins, P., Goerg, S., & Kube, S. (2017). *Self-chosen goals, incentives, and effort*. Mimeo.

Bruine de Bruin, W. B., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality Social Psychology, 92*, 342–350. https://doi.org/10.1037/0022-3514.92.5.938

Burns, D. J. (2007). An examination of the accuracy of students' expected grades. *Academy of Educational Leadersh Journal, 11*(3), 45–58.

Campitelli, G., & Labollita, M. (2010). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making, 5*(3), 182–191.

Caplan, D., Mortenson, K. G., & Lester, M. (2018). Can incentives mitigate student overconfidence at grade forecasts? *Accounting Education, 27*(1), 27–47. https://doi.org/10.1080/09639284.2017.1361850

Chen, P. P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning Indivdual Differences, 14*(1), 77–90. https://doi.org/10.1016/J.LINDIF.2003.08.003

Clark, D., Gill, D., Prowse, V., & Rush, M. (2020). Using goals to motivate college students: Theory and evidence from field experiments. *Review of Economics and Statistics, 102*(4), 648–663. https://doi.org/10.1162/rest_a_00864

Corgnet, B., Gómez-Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making, 4*(1), 20–33.

Colom, R., Escorial, S., Shih, P. C., & Privado, J. (2007). Fluid intelligence, memory span, and temperament difficulties predict academic performance of young adolescents. *Personality and Individual Differences, 42*(8), 1503–1514. https://doi.org/10.1016/J.PAID.2006.10.023

Cordero, A., & Corral, S. (2006). *DAT - 5 tests de aptitudes diferenciales*. TEA Ediciones Ed.

Corgnet, B., Gómez-Miñambres, J., & Hernán-González, R. (2015). Goal setting and monetary incentives: When large stakes are not enough. *Management Science, 61*(12), 2926–2944. https://doi.org/10.1287/mnsc.2014.2068

Croson, R. (2005). The Method of Experimental Economics. *International Negotiation, 10*(1), 131–148. https://doi.org/10.1163/1571806054741100

Dalton, P. S., Jimenez, V. H. G., & Noussair, C. (2015). Paying with self-chosen goals: Incentives and gender differences. Discussion Paper No 2015-021, Center for Economic Research. https://EconPapers.repec.org/RePEc:tiu:tiucen:35daceab-34bc-4bd2-b330-ec03da6b06b7

Diecidue, E., & van de Ven, J. (2008). Aspiration level, Probability of Success and Failure, and Expected Utility. *International Economic Review, 49*(2), 683–700. https://doi.org/10.1111/j.1468-2354.2008.00494.x

Dhall, S., & Thukral, P. (2009). Intelligence as Related to Self-confidence and Academic Achievement of School Students. *Journal of All India Association Education Research, 21*(2), 80–83.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review, 100*(3), 1238–1260. https://doi.org/10.1257/aer.100.3.1238

Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. *Advances in Experimental Social Psychology, 44*, 247–296. https://doi.org/10.1016/B978-0-12-385522-0.00005-6

Duttle, K. (2016). Cognitive skills and confidence: Interrelations with overestimation, overplacement and overprecision. *Bulletin of Economic Research, 68*(S1), 42–55. https://doi.org/10.1111/boer.12069

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes, 105*(1), 98–121. https://doi.org/10.1016/j.obhdp.2007.05.002

Falk, A., & Knell, M. (2004). Choosing the joneses: endogenous goals and reference standards. *Scandinavian Journal of Economics, 106*(3), 417–435. https://doi.org/10.1111/j.0347-0520.2004.00370.x

Feld, J., Sauermann, J., & de Grip, A. (2017). Estimating the relationship between skill and overconfidence. *Journal of Behavioral Experimental Economics, 68*, 18–24. https://doi.org/10.1016/J.SOCEC.2017.03.002

Foster, N. L., Was, C. A., Dunlosky, J., & Isaacson, R. M. (2017). Even after thirteen class exams, students are still overconfident: The role of memory for past exam performance in student predictions. *Metacognition Learn, 12*(1), 1–19. https://doi.org/10.1007/s11409-016-9158-6

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspective, 19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Fryer, J. W., & Elliot, A. J. (2008). *Self-regulation of achievement goal pursuit, motivation and self-regulated learning. Motivation and self-regulated learning: Theory, research, and applications* (p. 416). Lawrence Erlbaum Associates.

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition, 15*(1), 84–93. https://doi.org/10.3758/BF03197714

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspective, 25*(4), 191–210. https://doi.org/10.1257/JEP.25.4.191

Gómez-Miñambres, J. (2012). Motivation through goal setting. *Journal of Economic Psychology, 33*(6), 1223–1239. https://doi.org/10.1016/j.joep.2012.08.010

Grimes, P. W. (2002). The overconfident principles of economics student: An examination of a metacognitive skill. *The Journal Economic Education, 33*(1), 15–30.

Gutierrez, A. P., & Schraw, G. (2015). Effects of strategy training and incentives on students' performance, confidence, and calibration. *The Journal of Experimental Education, 83*(3), 386–404. https://doi.org/10.1080/00220973.2014.907230

Hack, A., & Von Bieberstein, F. (2015). How expectations affect reference point formation: An experimental investigation. *Review of Managerial Science, 9*(1), 33–59. https://doi.org/10.1007/s11846-014-0121-0

Hack, A., von Bieberstein, F., & Kraiczy, N. D. (2016). Reference point formation and new venture creation. *Small Business Economics.* https://doi.org/10.1007/s11187-015-9694-5

Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition Learn, 3*(2), 101–121. https://doi.org/10.1007/s11409-008-9021-5

Heath, C., Larrick, R. P., & Wu, G. (1999). Goals as reference points. *Cognitive Psychology, 38*(1), 79–109. https://doi.org/10.1006/cogp.1998.0708

Herranz-Zarzoso, N., & Sabater-Grande, G. (2018). Monetary incentives and self-chosen goals in academic performance: An experimental study. *International Review of Economics Education, 27*, 34–44. https://doi.org/10.1016/J.IREE.2018.02.002

Hoelzl, E., & Rustichini, A. (2005). Overconfident: Do you put your money on it? *The Economic Journal, 115*(503), 305–318. https://doi.org/10.1111/j.1468-0297.2005.00990.x

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review, 92*(5), 1644–1655. https://doi.org/10.1257/000282802762024700

Hoppe, E. I., & Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. *Economic Letters, 110*(2), 97–100. https://doi.org/10.1016/J.ECONLET.2010.11.015

Hossain, B., & Tsigaris, P. (2013). Can business students forecast their own grade? *American Journal of Business Education, 6*(1), 85–92.

Hossain, B., & Tsigaris, P. (2015). Are grade expectations rational? *A Classroom Experiment Education Economics, 23*(2), 199–212. https://doi.org/10.1080/09645292.2012.735073

Jensen, P. A., & Moore, R. (2008). Students' behaviors, grades and perceptions in an introductory biology course. *The American Biology Teacher, 70*(8), 483–487. https://doi.org/10.1662/0002-7685(2008)70[483:SBGPIA]2.0.CO;2

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–292. https://doi.org/10.1142/9789814417358_0006

Khachikian, C. S., Guillaume, D. W., & Pham, T. K. (2011). Changes in student effort and grade expectation in the course of a term. *European Journal of Engineering Education, 36*(6), 595–605. https://doi.org/10.1080/03043797.2011.637189

Koszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics, 121*(4), 1133–1165. https://doi.org/10.1093/qje/121.4.1133

Koszegi, B., & Rabin, M. (2007). Reference-dependent risk attitudes. *American Economic Review.* https://doi.org/10.1257/aer.97.4.1047

Koszegi, B., & Rabin, M. (2009). Reference-dependent consumption plans. *American Economic Review, 99*(3), 909–936. https://doi.org/10.1257/aer.99.3.909

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personalityand Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Lavecchia, A. M., Liu, H., & Oreopoulos, P. (2016). Behavioral economics of education: Progress and possibilities. *Handbook of the Economic Education, 5*, 1–74. https://doi.org/10.1016/B978-0-444-63459-7.00001-4

List, J. A., Livingston, J. A., & Neckermann, S. (2018). Do financial incentives crowd out intrinsic motivation to perform on standardized tests? *Economics of Education Review, 66*, 125–136. https://doi.org/10.1016/J.ECONEDUREV.2018.08.002

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705–717. https://doi.org/10.1037/0003-066X.57.9.705

Magnus, J. R., & Peresetsky, A. A. (2018). Grade expectations: Rationality and overconfidence. *Frontiers in Psychology, 8*, 2346. https://doi.org/10.3389/fpsyg.2017.02346

Maki, R. H. (1998). Test predictions over text material. In Erlbaum (Ed.), *Metacognition in educational theory and practice* (pp. 117–144). Routledge. https://doi.org/10.4324/9781410602350-13

Markle, A., Wu, G., White, R., & Sackett, A. (2018). Goals as reference points in marathon running: A novel test of reference dependence. *Journal of Risk and Uncertainty, 56*(1), 19–50. https://doi.org/10.1007/s11166-018-9271-9

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1–10. https://doi.org/10.1016/J.INTELL.2008.08.004

Meloy, M. G., Russo, J. E., & Miller, E. G. (2006). Monetary incentives and mood. *Journal of Marketing Research, 43*(2), 267–275. https://doi.org/10.1509/jmkr.43.2.267

Meng, J. (2019). To gain with less pain?: Reference-dependence in students' exam performance and effort provision. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3505051

Miller, T. M., & Geraci, L. (2011). Training metacognition in the classroom: The influence of incentives and feedback on exam predictions. *Metacognition Learn, 6*(3), 303–314.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502–517. https://doi.org/10.1037/0033-295X.115.2.502

Murstein, B. I. (1965). The relationship of grade expectations and grades believed to be deserved to actual grades received. *The Journal of Experimental Education, 33*(4), 357–362. https://doi.org/10.1080/00220973.1965.11010894

Nicholson, L., Putwain, D., Connors, L., & Hornby-Atkinson, P. (2013). The key to successful achievement as an undergraduate student: Confidence and realistic expectations? *Studies in Higher Education, 38*(2), 285–298. https://doi.org/10.1080/03075079.2011.585710

Nowell, C., & Alston, R. M. (2007). I thought I got an A! overconfidence across the economics curriculum. *Journal of Economic Education, 38*(2), 131–142. https://doi.org/10.3200/JECE.38.2.131-142

Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization, 72*(1), 147–152. https://doi.org/10.1016/J.JEBO.2009.04.018

Park, Y. J., & Santos-Pinto, L. (2010). Overconfidence in tournaments: Evidence from the field. *Theory and Decision, 69*, 143–166. https://doi.org/10.1007/s11238-010-9200-0

Saenz, G. D., Geraci, L., & Tirso, R. (2019). Improving metacognition: A comparison of interventions. *Applied Cognitive Psychology, 33*(5), 918–929. https://doi.org/10.1002/acp.3556

Sander, P., De La Fuente Arias, J., Stevenson, K., & Jones, T. (2011). A validation of the academic behavioural confidence scale with spanish psychology students. *Psychology Learning and Teaching, 10*(1), 11–24. https://doi.org/10.2304/plat.2011.10.1.11

Sander, P., & Sanders, L. (2009). Measuring academic behavioural confidence: The ABC scale revisited. *Studies in Higher Education, 34*(1), 19–35. https://doi.org/10.1080/03075070802457058

Santos-Pinto, L., & de la Rosa, L. E. (2020) Overconfidence in labor markets. In Handbook of Labor, Human Resources and Population Economics, pp. 1–42.

Serra, M. J., & DeMarree, K. G. (2016). Unskilled and unaware in the classroom: College students' desired grades predict their biased grade predictions. *Memory and Cognition, 44*(7), 1127–1137. https://doi.org/10.3758/s13421-016-0624-9

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*(2), 161–188.

Sturges, D., Maurer, T. W., Allen, D., Gatch, D. B., & Shankar, P. (2016). Academic performance in human anatomy and physiology classes: A 2-yr study of academic motivation and grade expectation. *Advances in Physiology Education, 40*(1), 26–31. https://doi.org/10.1152/advan.00091.2015

Svanum, S., & Bigatti, S. (2006). Grade Expectations: Informed or uninformed optimism, or both? *Teaching of Psychology, 33*(1), 14–18. https://doi.org/10.1207/s15328023top3301_4

Tandon, R. K. (1971). Manual of Directions : The Group Test of General Mental Ability (A Point Scale for Adults). K.G.K. College Building, Muradabad

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297–323. https://doi.org/10.1007/BF00122574

van Lent, M., & Souverijn, M. (2020). Goal setting and raising the bar: A field experiment. *Journal of Behavioral and Experimental Economics, 87*, 101570. https://doi.org/10.1016/j.socec.2020.101570

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology, 27*(4), 551–572. https://doi.org/10.1016/S0361-476X(02)00006-1

Wolfe, R. N., & Grosch, J. W. (1990). Personality correlates of confidence in one's decisions. *Journal of Personality, 58*(3), 515–534. https://doi.org/10.1111/j.1467-6494.1990.tb00241.x

Wu, G., Heath, C., & Larrick, R. (2008). A prospect theory model of goal behavior. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.554.7677&rep=rep1&type=pdf (Unpublished manuscript)