



Improving accuracy in predictions about future performance

Joelle C. Ruthig¹ · Abigail M. Kroke²

Accepted: 16 April 2024
© The Author(s) 2024

Abstract

Within the academic context, students often expect to perform significantly better on an upcoming exam when compared to their actual performance. Unfortunately, students who are least accurate in their performance predictions are also the most at risk of being underprepared compared to their peers. Thus, efforts to enhance accuracy in performance predictions would benefit the student population that is most at risk of academic shortcomings and failure. Two studies examined whether incentives improve accuracy of academic performance predictions. In Study 1, 126 students in a 200-level undergraduate course provided estimates of how well they would perform on an upcoming exam. Two weeks later, the same students were randomly assigned to receive a reward incentive, punishment incentive, or no incentive for accuracy in performance predictions and their exam performance estimates were reassessed then compared to their actual scores on an exam completed one week later. In Study 2, 144 students from multiple 300-level courses provided performance predictions for an upcoming exam that were then compared to their actual exam scores for accuracy. For a subsequent exam, the same students were randomly assigned to varying levels of reward incentives (no incentive, modest incentive, large incentive) for prediction accuracy and their performance estimates were examined and compared to actual performance on a second exam. Findings from both studies indicated performance estimates are malleable and with the appropriate incentives, accuracy in predictions may be improved and contribute to better actual performance.

Keywords Unrealistic optimism · Performance · Prediction accuracy · College students

Improving accuracy in predictions about future performance

Unrealistic optimism is the expectation of personal outcomes that are more favorable than actually occur (Weinstein, 1980). This cognitive bias has been widely demonstrated in various contexts over the last several decades (Shepperd et al., 2013) and its consequences are mixed. Some findings suggest unrealistic optimism in the form of overestimating performance can enhance psychological well-being, reflecting hope for future achievement and reducing performance anxiety. Armor and Sackett (2006) indicate that overestimating future performance contributes to pride and self-esteem. In fact, some researchers suggest overly

optimistic performance predictions may become a self-fulfilling prophecy such that optimistically-biased expectations help to bring about expected outcomes (Armor & Taylor, 2002, 2003) by motivating individuals to achieve those lofty standards (Zhang & Fishbach, 2010).

Conversely, other research suggests negative consequences for unrealistic performance estimates. As Dunning et al. (2004) acknowledge, inaccuracy in personal performance estimates may entail consequences ranging from negative affect such as disappointment, guilt, and shame, to lost effort or missed opportunities for oneself, through fatal outcomes that extend beyond the individual (e.g., overestimating one's ability to drive a car full of passengers under conditions of poor visibility).

Within the academic context, unrealistic optimism may include overestimating how quickly an assignment will be completed (Buehler et al., 1994) or expecting to perform significantly better on an upcoming exam when compared to actual performance (Ruthig et al., 2017a). Although these examples of academic unrealistic optimism are unlikely to result in life-or-death consequences, students who are least

✉ Joelle C. Ruthig
joelle.ruthig@und.edu

¹ University of North Dakota, 501 N. Columbia Rd,
58202 Grand Forks, ND, USA

² University of Illinois-Springfield, Springfield, IL, USA

accurate in their performance estimates are also the most at risk of being underprepared compared to their peers who have greater self-awareness and more accurate performance predictions (Jansen et al., 2021; Simons, 2013). This phenomenon, referred to as the Dunning-Kruger effect (Kruger & Dunning, 1999), indicates that unrealistic optimism may be most prominent among the weakest students. Thus, efforts to enhance accuracy in performance predictions would benefit the student population that is most at risk of academic shortcomings and failure.

Assessing and incentivizing performance prediction accuracy

Greater realism about performance may promote preparedness, motivation, and better predictability. Thus, it is vital to examine accuracy in performance predictions and how accuracy may be enhanced to minimize risk of academic failure and attrition. However, as noted by Armor and Sackett (2006), prior research has been somewhat limited in determining accuracy in predictions because participants' performance outcomes could often not be assessed. Instead, most research on unrealistic optimism and performance estimates has focused on anonymous, group-level predictions for future events (e.g., students rating the likelihood of landing a successful career) lacking verification of individual-level outcomes (Jefferson et al., 2017).

In studies that were able to verify individuals' performance outcomes, limited research has systematically examined interventions aimed at improving performance prediction accuracy, or provided adequate incentives for prediction accuracy. For example, Hacker et al. (2008) offered bonus points for greater accuracy among college students in estimating their course exam performance. Hacker et al. found that such incentive improved prediction accuracy among a sub-set of lower-performing students. In other research that has provided incentives in an attempt to enhance accuracy, the incentives may have been insufficient. For example, Lewine and Sommers (2016) assessed students' accuracy in predicting their performance immediately prior to an upcoming exam, offering students up to three bonus points on the exam for accuracy. Lewine and Sommers found minimal differences in prediction accuracy of the incentivized group compared to a non-incentivized control group. These findings suggest that the incentive was not sufficient to motivate accuracy, however, the study did not assess the impact of magnitude of incentive on prediction accuracy.

In their research, Simmons and Massey (2012) did examine the impact of magnitude of incentive on prediction accuracy. They compared a small monetary incentive (\$5)

to a large monetary incentive (\$50) for accuracy in predicting a performance outcome: which NFL team would win a football game. These researchers found overly optimistic predictions prevailed even in the large monetary incentive condition, suggesting incentives had limited impact on accuracy in prediction (Simmons & Massey, 2012). However, performance predictions in this study focused on an outcome that was outside of the predictor's control (i.e., which sports team would win). Consequently, these findings do not provide insight into the impact of magnitude of incentive on performance accuracy when the predictor has control over the performance event. Although understanding accuracy in predictions for events outside of one's control may contribute to understanding the boundaries of unrealistic optimism, it is critical to assess accuracy in performance predictions involving performance that is within one's personal control. Moreover, such predictions should involve real performance in the near future, not hypothetical or distant future performance predictions (Li et al., 2021), nor be combined with post-diction estimates (Hacker et al., 2008) that cannot retrospectively impact the performance of a recently completed event.

Collectively, prior research on incentivizing performance estimate accuracy is often limited in terms of focusing on performance outside of one's control (e.g., Simmons & Massey, 2012), examining performance that has relatively low stakes (e.g., scavenger hunt performance; Amor & Sackett, 2006), and assessing only a single incentive to motivate prediction accuracy (Lewine & Sommers, 2016). As such, it is uncertain whether unrealistically optimistic performance estimates are resistant to change via incentives or whether the optimal magnitude of incentive has yet to be identified. Consequently, potential benefits of enhanced accuracy in future performance predictions remain largely unknown.

The current studies systematically tested whether incentives improve accuracy of performance predictions in academic contexts. This will clarify whether unrealistically optimistic estimates are malleable, given appropriate incentives, and whether improved accuracy in performance predictions relates to actual performance. Together, these studies will advance research on unrealistic optimism and performance predictions while providing insight into performance within the academic domain.

Study 1

Study 1 involved three phases spanning several weeks, focused on an undergraduate course exam as the performance event, and included actual performance outcomes

(i.e., exam scores). Performance estimates for an upcoming exam were assessed one month prior to the exam, then again one week before the exam. Prior to the second performance prediction, students were randomly assigned to receive either a reward for a given level of accuracy in their prediction, punishment for failing to achieve a certain level of accuracy, or no incentive for prediction accuracy. This design enabled us to identify individual-level accuracy of performance estimates. We were also able to examine both within-group changes in accuracy as a result of incentive exposure, and between-group differences in accuracy as a function of type of incentive exposure. Finally, the study design allowed us to determine whether greater prediction accuracy was associated with better actual performance.

The first objective of Study 1 was to assess initial academic performance estimates for an upcoming course exam without any incentive introduced. As in past research (Ruthig et al., 2017a; Shepperd et al., 1996), we expected initial performance predictions to be overestimated or unrealistically optimistic compared to actual performance.

The second objective was to determine whether accuracy incentives would result in greater accuracy in performance predictions, as well as improvement in accuracy of performance predictions when compared to earlier estimates. One week prior to the exam, students were asked to estimate their exam score out of 100% on the upcoming exam and those randomly assigned to the reward condition were informed that if their estimate was within 5% of their actual exam score, they would receive two bonus points on that exam. This 5% accuracy threshold is consistent with prior research (e.g., Zhang & Fishbach, 2010). Students in the punishment condition were told that if their estimate did not fall within 5% of their actual exam score, their instructor would deduct two points from their exam score. The control group simply provided an estimate with no incentive. This allowed for direct comparison of accuracy incentive conditions, as well as assessment of changes in individual-level performance prediction accuracy. We expected the incentive conditions to result in greater accuracy: students who received the reward or punishment incentive would be more accurate in predicting their upcoming exam performance than those who received the control/no incentive condition. We also expected that students who received an incentive (reward or punishment) would improve in prediction accuracy from their first estimate prior to incentive to their second estimate after the incentive was introduced. No such improvement was expected among the control group.

The third objective was to determine the association between accuracy in predictions and actual performance.

Based on the Dunning-Kruger effect (Kruger & Dunning, 1999), we expected a positive association between prediction accuracy and performance, such that students who are more accurate in their performance predictions will outperform students who are less accurate in their performance predictions.

Method

Participants

Upon receiving approval of the study from the IRB at the University of North Dakota, undergraduate students in a 200-level psychology course were invited to participate in the study for two extra credit points toward their course. All 126 students enrolled in the course chose to participate and the majority (74.4%) identified as female ($n = 96$), another 23.3% of participants identified as male ($n = 30$), and the remaining 2.3% of participants did not indicate their gender ($n = 3$). Participants ranged in age from 18 to 26 years old, with an average age of 19.33 ($SD = 1.53$), and primarily identified their race as White (87%; $n = 112$). All college levels were represented with 60 freshmen, 52 sophomores, 7 juniors, and 7 seniors from 22 different undergraduate programs, with the majority from Nursing (31%; $n = 40$), Psychology (17%; $n = 22$), Biology (11.6% $n = 15$), and Physical Therapy (8.5%; $n = 11$).

Measures

Accuracy incentives All participants were told: “*Accuracy in estimating future performance is important.*” Those participants randomly assigned to either the reward or punishment accuracy incentive then received the following verbal instructions: “*Your course instructor has agreed to provide incentive for accurately estimating your grade for the upcoming exam.*” Participants in the reward incentive condition were then told: “*If your exam grade estimate is within 5% of your actual exam score, your instructor will add 2 bonus points to your exam score.*” Participants in the punishment incentive condition were told: “*If your exam grade estimate is NOT within 5% of your actual exam score, your instructor will deduct 2 points from your exam score.*”

Performance estimates As in prior research on academic performance estimates (e.g., Ruthig et al., 2022), participants were asked to provide an estimate ranging from 0% to

100% in response to the following: “My grade on the exam in this course will be ___%”.

Performance outcomes With students’ consent, their course instructor provided their exam scores to the researchers as percentages out of 100%.

Procedure

During a class period one month prior to an upcoming exam, participants completed a brief survey asking about their performance expectations for that exam. Next, at the start of a class period one week before the exam, participants responded to another brief survey asking about their performance expectations for the upcoming exam. Prior to providing their performance expectations, they received verbal instructions about accuracy based on which one of the three incentive conditions they were assigned to: control/no incentive, reward incentive, or punishment incentive. Immediately after participants submitted their exam estimates, they were fully debriefed that there were no actual points being added or deducted from their exam score for prediction accuracy. One week later, participants completed their exam and the course instructor subsequently provided their actual exam scores to the researchers.

Results

The first study objective was to assess accuracy of academic performance estimates before any incentive was introduced. As expected, a paired-samples t-test comparing students’ performance estimates one month before the exam to their actual exam score indicated that they significantly overestimated how well they would perform, $t(112) = 4.76, p = .001, d = 0.45$. On average, participants predicted they would score 87.84% on the exam ($SD = 6.67$), whereas they only scored an average of 84.50% ($SD = 7.98$), overestimating their upcoming performance by more than 3%.

For the second objective, one week prior to the exam, participants were randomly assigned to a reward incentive, a punishment incentive, or no incentive/control condition, to determine whether incentives would result in greater accuracy in performance predictions, as well as improvement in accuracy performance predictions when compared to earlier predictions. Of the 126 participants who completed performance estimates a month prior to the exam, seven participants were absent from class and missed the incentive manipulation, resulting in 119 participants who completed the remainder of the study. Of those 119 participants, 38 (32%) received the control/no incentive condition,

46 (38.5%) received the reward incentive condition, and 35 (29.5%) received the punishment condition.¹

We then identified how many participants within each of the three incentive conditions, provided performance estimates that fell within 5% of their actual exam score. Among the 38 participants assigned to the no incentive condition, 22 (57.9%) predicted their exam score within 5% accuracy. Similarly, 20 of 35 participants (57%) in the punishment incentive condition predicted their exam score within 5% accuracy. Of the 46 participants who received the reward incentive, 30 or 65% predicted their exam score within 5%. Thus, regardless of incentive condition, more students predicted their performance within the 5% accuracy threshold, with the reward incentive condition eliciting the greatest proportion within that 5% accuracy threshold.

Next, we computed an accuracy difference score (i.e., each student’s performance estimate minus their actual score), such that positive values indicate overestimation, negative values indicate underestimation, and scores closer to 0 reflect greater accuracy. A one-way Analysis of Variance (ANOVA) with the accuracy difference score as the dependent measure, was computed to assess between-group differences in accuracy of performance estimates as a function of incentive condition. Although the reward incentive group appeared to be the most accurate, overestimating their performance to a lesser degree ($M = 1.91, SD = 5.97$) than the no incentive/control group ($M = 3.18, SD = 6.96$) or punishment incentive group ($M = 3.57, SD = 10.67$), there were no significant between-group differences in performance estimate accuracy, $F(2,116) = 0.50, p = .61$.

After assessing between-group differences, we examined whether there were within-group improvements in prediction accuracy as a function of accuracy incentive. Paired-samples t-tests were used to assess whether there were within-group changes in prediction accuracy from pre-incentive to post-incentive estimates. As detailed in Table 1, results showed significant improvement in prediction accuracy among the punishment incentive group, with their post-incentive performance estimate ($M = 2.52$) being significantly more accurate than their pre-incentive performance estimate ($M = 3.83$).

The final Study 1 objective was to assess the association between accuracy in predictions and actual performance. Collapsing across incentive conditions and controlling for college level, a one-way ANOVA was computed to compare students whose performance prediction one week before the exam was within 5% of their actual score to those whose performance prediction fell beyond 5% accuracy, with their

¹ Retrospectively, to ensure no pre-incentive prediction differences existed between these groups, we computed a one-way ANOVA comparing their earlier performance predictions (one month prior to the exam) and found no significant group differences: $F(2,109) = 0.81, ns$.

Table 1 Study 1: Within-group changes in prediction accuracy as a function of incentive

Incentive type	Pre-incentive accuracy		Post-incentive accuracy		t (df)	95% CI L U	One-tailed <i>p</i>
	M	(SD)	M	(SD)			
No incentive	3.84	(8.06)	3.18	(6.96)	0.86 (37)	-0.89, 2.21	0.198
Reward	2.55	(6.94)	2.14	(5.92)	0.61 (43)	-0.95, 1.77	0.273
Punishment	3.83	(7.82)	2.52	(6.13)	2.05 (28)	0.01, 2.62	0.025

Mean values closer to zero reflect greater accuracy

actual performance as the dependent measure. As expected, students who were within 5% accuracy in their performance prediction significantly outperformed those whose performance prediction was beyond 5% accuracy ($M=86.97$, $SD=5.86$ vs. $M=79.52$, $SD=11.41$), $F(1, 116)=21.80$, $p=.000$, $\eta_p^2=0.158$.

Discussion

Study 1 findings were consistent with prior research (e.g., Ruthig et al., 2017b; Shepperd et al., 1996) showing that early predictions about one's future academic performance a month away were significantly overestimated. That is, prior to an offer of incentive for accuracy (or disincentive for inaccuracy), students anticipated achieving an exam score that was approximately 3% higher than their actual exam grade.

Regarding improvement in prediction accuracy from pre- to post-accuracy incentive, only the punishment incentive yielded significant improvement. Students informed that they would lose points on the exam, unless their performance prediction fell within 5% of their actual exam grade, were significantly more accurate in predicting their exam score compared to earlier estimates before the threat of punishment for inaccuracy. Likewise, students who were in the reward incentive condition and informed that they were eligible for bonus points if their performance estimates were within 5% accuracy also appeared to improve in their prediction accuracy, although the improvement did not reach statistical significance. In contrast, students who did not receive any incentive appeared to become less accurate in their subsequent prediction, though not significantly so. Together, these finds provide tentative support for the notion that incentivizing accuracy in performance predictions may contribute to more accurate estimates within the academic context.

Study 1 results also support the Dunning-Kruger effect (Kruger & Dunning, 1999). Specifically, students who predicted their score within 5% of their actual exam score significantly outperformed those students whose estimates were beyond 5% accuracy. Thus, accuracy in performance predictions matters – greater accuracy in estimates one week prior to the performance event was associated with

significantly better exam performance. The weakest students were the most optimistic and most inaccurate.

Although all three accuracy incentive groups had greater frequency of estimating within (vs. outside) 5% of their actual score, the reward incentive group had a greater proportion of estimates within 5% of their actual score. However, these preliminary findings are limited in terms of applying accuracy incentives only within a single course in the academic domain. We attempted to address this and other limitations of Study 1 in a subsequent study as described below.

Study 2

In a second study on prediction accuracy in performance estimates, we attempted to build upon Study 1 in four important ways. First, Study 1 focused on a single college course whereas Study 2 included performance predictions across three separate 300-level college courses. In addition, Study 1 examined only a single performance event. In Study 2, we assessed two performance events: Exam 1 and Exam 2. Adding a second performance event allowed us to establish a baseline performance prediction and event outcome prior to introducing the accuracy incentives. Third, Study 1 included a single reward incentive, whereas in Study 2 we assessed different magnitudes of reward incentives. Finally, in Study 2 we examined explanations for performance estimates in order to identify the most common factors that participants based their exam performance estimates on.

Similar to Study 1, students who participated in Study 2 provided performance estimates for an upcoming exam one week before the first exam then those estimates were compared to students' actual exam scores. Next, one week prior to Exam 2, we introduced an experimental manipulation prior to the Exam 2 performance prediction in which students were randomly assigned to receive a small reward for a given level of accuracy in their prediction, a larger reward for a given level of accuracy, or no incentive for prediction accuracy. This permitted us to determine individual-level accuracy of performance predictions, and track both within-group changes in accuracy following incentive exposure and between-group differences in accuracy as a function of magnitude of incentive exposure. Finally, as in Study 1, we

assessed the relationship between accuracy in performance predictions and actual performance.

Consistent with Study 1, the first objective of Study 2 was to assess performance estimates for an upcoming course exam (Exam 1) without any incentive introduced. We then compared those estimates to actual Exam 1 performance to assess individual-level accuracy in performance predictions. We again expected initial performance estimates to be unrealistically optimistic when compared to actual Exam 1 performance.

As in Study 1, Study 2's second objective was to assess whether incentives would lead to more accuracy and improvement in accuracy of performance predictions. To address this objective, one week prior to Exam 2, students were randomly assigned to a small reward incentive, large reward incentive, or no incentive/control condition. All students were again asked to predict their exam score out of 100% and those in the small reward condition were informed that if their estimate was within 5% of their actual exam score, they would receive two bonus points on that exam. This level of reward incentive is consistent with that used in Study 1. Participants in the large reward condition were told that if their estimate was within 5% of their actual exam score, they would receive 5 bonus points on Exam 2. The control group simply provided an estimate with no incentive. This allowed us to compare magnitude of accuracy incentive conditions to determine which is more effective if any, as well as directly examine changes in individual-level performance prediction accuracy. We expected the larger incentive condition to result in greater accuracy than the small incentive condition, which in turn, should result in greater accuracy than the no incentive condition. We also expected that students who received an incentive (either small or large reward) would improve in prediction accuracy from their first estimate prior to incentive to their second estimate after the incentive was introduced. No improvement was anticipated for the control group.

The third objective of Study 2 was the same as in Study 1: to determine the association between accuracy in predictions and actual performance. Consistent with the Dunning-Kruger effect (Kruger & Dunning, 1999), we again expected to find a positive relationship between prediction accuracy and performance, with students making the least accurate performance predictions also demonstrating the poorest actual performance.

The final Study 2 objective was exploratory in nature, namely to examine the reasons provided for performance estimates in order to identify any emerging themes. As Foster et al. (2017; p. 14) noted, insight into the source of students' exam performance estimates "may ultimately inform how to improve overall student achievement."

Method

Participants

Participants in the current study were 144 undergraduates enrolled in one of three 300-level psychology courses who completed the study during the first few minutes of class time in exchange for extra research credit. The sample self-identified as predominantly White ($n=125$, 86.8%), with the remaining participants identifying as Native American ($n=7$; 4.9%), Hispanic American ($n=6$; 4.2%), African American ($n=3$; 2.1%), or Asian American ($n=3$; 2.1%). Participants identified as male ($n=23$, 16%), female ($n=120$; 83.3%), or transgender ($n=1$, 0.7%) and were an average of 20.60 years old ($SD=1.83$). They indicated their college level as sophomore ($n=21$; 14.6%), junior ($n=72$; 50%), or senior ($n=51$, 35.4%).

Of the 144 initial participants who completed performance estimates for Exam 1, 42 participants were absent from class due to Covid-19, other illness, or other reasons (e.g., athletic travel) and missed the incentive manipulation for Exam 2, resulting in 102 participants who completed the remainder of the study. Of those 102 participants, 37 (36.3%) received the control/no incentive condition, 31 (30.4%) received the 2-point reward incentive condition, and 34 (33.3%) received the 5-point reward incentive condition.

Measures

Accuracy incentive Immediately prior to estimating their Exam 2 performance, participants were exposed to the control/no incentive condition or an incentive condition (2 or 5 bonus points) based on random assignment: "*Accuracy in estimating future performance is important. Your course instructor has agreed to provide incentive for accurately estimating your grade for Exam 2. If your Exam 2 grade estimate is within 5% of your actual Exam 2 score, your instructor will add 2 (or 5) bonus points to your exam score.*"

Control/no incentive condition: "*Accuracy in estimating future performance is important.*"

Performance estimates Participants provided estimates of their first and second exam performance by completing the following statement: "My grade on Exam 1 (or Exam 2) in this course will be ___%."

We also asked participants what their performance estimate was based on for each exam. They were asked the following open-ended question: "*What is the MAIN reason for the*

exam estimate you provided? That is, what is the MAIN factor you are basing the estimate on?"

Exam scores Course instructors provided the researchers with each student's actual score as a percentage out of a possible 100% for both exams.

Procedure

The study, approved by the University of North Dakota IRB, consisted of two phases of data collection, based around two exams administered during the fall semester. The course instructors agreed to give students enrolled in their course the opportunity to take part in this study in exchange for extra course credit. At the start of the semester, students were informed of the study and the opportunity to participate. Those who agreed to participate completed two brief surveys at two points in the semester.

Time 1 took place during a class period one week before Exam 1. Participants completed a brief survey about their expectations for Exam 1 along with various demographic measures. With participants' consent, their Exam 1 scores were collected from their course instructors.

Time 2 occurred during a class period one week before Exam 2 was administered, with participants responding to a brief survey asking about their expectations for Exam 2. Prior to providing their Exam 2 expectations, they read one of the randomly assigned "accuracy incentive" information statements. Immediately after participants submitted their Exam 2 estimates, they were told there were no actual points being added for accuracy and were fully debriefed as to the purpose of the study. Again, with their consent, participants' Exam 2 scores were collected from their course instructors.

Results

Participants' Exam 1 performance estimates and actual exam scores were compared across accuracy incentive groups to ensure there were no pre-exposure to accuracy incentive differences. The overall MANCOVA, with college level as a covariate, indicated no significant baseline differences in Exam 1 performance estimates or actual performance: Wilks lambda = 0.997, $F(4, 196) = 0.18$, $p = .948$. Specifically, as detailed in Table 2, all three groups estimated that they would score approximately 87% on Exam

1, and all three groups actually scored an average of 84% on Exam 1. Similarly, a repeated measures ANCOVA, controlling for college level, indicated that groups did not differ in the degree to which they overestimated their Exam 1 score: $F(2, 99) = 0.28$, $p = .76$. Thus, students tended to overestimate their Exam 1 performance and no pre-manipulation differences were found in estimates or actual performance for Exam 1.

Having established that there were no baseline differences in performance estimates, prediction accuracy or actual performance, we shifted to the main objective of assessing whether providing incentives of varying magnitudes impacted accuracy in performance estimates. As in Study 1, we first assessed how many participants within each of the three incentive conditions provided performance estimates that fell within 5% of their actual exam score. Among the 39 participants assigned to the no incentive condition, 20 (51%) predicted their exam score within 5% accuracy. Of the 40 participants who received the two-point reward incentive, 25 (62.5%) predicted their exam score within 5% accuracy. Finally, of the 42 participants who received the five-point reward incentive, 23 or 55% predicted their exam score within 5%. Thus, it appears the two-point reward incentive condition elicited the greatest proportion within that 5% accuracy threshold.

Next, consistent with Study 1, we assessed the association between accuracy in Exam 2 predictions and actual Exam 2 performance. Collapsing across incentive conditions and controlling for college level, a one-way ANOVA was computed to compare students whose performance predictions one week before Exam 2 were within 5% of their actual score to those whose performance predictions fell beyond 5% accuracy, with their actual performance as the dependent measure. As expected and consistent with Study 1, students in Study 2 who were within 5% accuracy in their performance prediction for Exam 2 significantly outperformed those whose performance predictions were beyond 5% accuracy ($M = 89.14$, $SD = 5.63$ vs. $M = 81.32$, $SD = 10.04$), $F(1, 109) = 28.38$, $p = .000$, $\eta_p^2 = 0.207$.

To compare prediction accuracy between groups, a univariate ANCOVA with college level as a covariate was computed using a difference score as the dependent variable (estimate minus actual score such that positive scores indicated overestimation, negative scores indicated underestimation, with greater departure from 0 in either direction indicating less accuracy). As depicted in Fig. 1, results

Table 2 Study 2: Baseline (pre-incentive exposure) group comparisons of predicted Exam 1 performance and actual Exam 1 score

Group	Predicted performance		Actual performance		Amount of inaccuracy	
	M	(SD)	M	(SD)	M	(SD)
No Incentive	87.32%	(3.94)	84.70%	(9.73)	2.62%	(8.59)
2 Point Incentive	87.85%	(4.64)	84.30%	(7.86)	3.55%	(7.59)
5 Point Incentive	87.36%	(6.05)	84.42%	(8.13)	2.94%	(8.11)

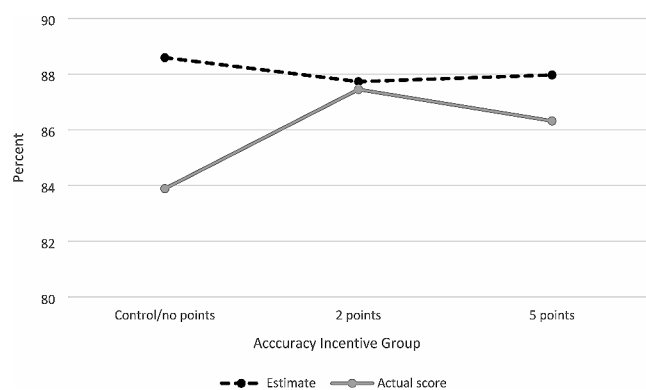


Fig. 1 Study 2: Accuracy in predicting Exam 2 performance as a function of magnitude of incentive

indicated a significant effect for incentive on performance accuracy: $F(2, 198) = 3.58, p = .032$. Specifically, the control/no incentive group was the least accurate ($M = 4.81; SD = 8.50$), and planned contrasts indicated that the 2-point incentive group was significantly more accurate ($M = 0.27; SD = 6.07$) than the control/no incentive group: $t = -4.78, 95\% CI: -8.43, -1.13, p = .011$. The 5-point incentive group's prediction accuracy ($M = 1.65; SD = 7.77$) was not significantly different from the control group ($M = 4.81; SD = 8.50$); $t = -3.34, 95\% CI: -6.97, 0.30, p = .071$. Thus, providing a 2-point incentive resulted in significantly greater prediction accuracy than the control/no incentive condition.

Next, we computed simple paired-samples t-tests to assess whether there were within-group changes in prediction accuracy from Exam 1 (pre-incentive) to Exam 2 (post-incentive). As detailed in Table 3, results showed significant improvement in prediction accuracy among the 2-point incentive group only. Although not statistically significant, the control/no-incentive group's accuracy in predicting their exam performance trended towards worsening from Exam 1 to Exam 2.

Another objective was to determine whether any of the groups improved in their actual performance from Exam 1 to Exam 2. As illustrated in Fig. 2, only the 2-point incentive group showed a significant improvement in actual performance from Exam 1 ($M = 84.30, SD = 7.86$) to Exam 2 ($M = 87.45, SD = 7.45$), $t(32) = -2.82, 95\% CI -5.43, -0.88, p = .008$. Although the 5-point incentive group appeared to improve in their performance Exam 1 ($M = 84.42, SD = 8.13$) to Exam 2 ($M = 86.32, SD = 8.49$), the change was not statistically significant: $t(32) = -1.19, 95\% CI -5.13, 1.34, p = .242$. The control/no-incentive group showed no sign

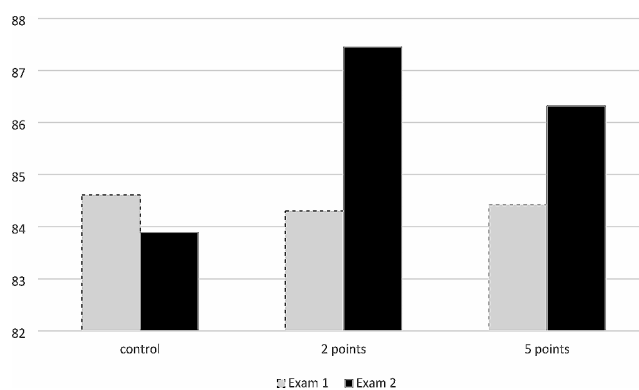


Fig. 2 Study 2: Comparison of actual performance change from Exam 1 to Exam 2 as a function of magnitude of incentive

of improvement in performance from Exam 1 ($M = 84.61, SD = 9.85$) to Exam 2 ($M = 83.89, SD = 9.41$), $t(35) = 0.53, 95\% CI -1.88, 3.33, p = .57$.

As a final objective, we examined participants' explanations for their performance estimates. Open-ended responses to "What is the main reason for your performance estimate?" for both exams were entered into Dedoose (Version 9.0.17; 2021) for a conceptual content analysis using inductive coding to reveal themes. As detailed in Fig. 3, a total of 22 codes were generated. Inter-rater agreement was based on pooled Cohen's kappa to capture inter-rater reliability across multiple codes (De Vries et al., 2008). The resulting inter-rater reliability was $\kappa = 0.71$, indicating "good agreement" between raters (Cicchetti, 1994; Landis & Koch, 1977).

For Exam 1, participants' most common explanation for their performance estimate was preparation behaviors (52 respondents; sample excerpt: "I have attended most lectures, studied notes, and read the chapters"). The second most common explanation was comprehension of course material (45 respondents; sample excerpt: "I understand the course material well"), followed by acknowledging that it was the first exam in the course, presumably making it difficult to know what the exam would be like (30 respondents; sample excerpt: "It is the first test in the course so I don't know what to expect").

For Exam 2, the most common reason participants gave for their performance estimates was previous exam performance (62 respondents; sample excerpt: "based on my last test score in this course") and then comprehension of material (25 respondents; sample excerpt provided above) and then what participants referred to as their confidence level

Table 3 Study 2: Within-group changes in prediction accuracy as a function of incentive

Group	Exam 1 accuracy		Exam 2 accuracy		t (df)	95% CI L U	One-tailed p
	M	(SD)	M	(SD)			
No incentive	2.92	(8.52)	4.81	(8.50)	-1.06 (35)	-5.49, 1.71	0.147
2 Point Incentive	3.55	(7.59)	0.27	(6.07)	2.09 (32)	0.08, 6.47	0.022
5 Point Incentive	2.94	(8.27)	1.65	(7.77)	0.61 (32)	-2.99, 5.56	0.272

Media	Codes																				Totals	Totals	
	Academic Workload	Avoiding disappointment	Class Attendance	Comprehension of Material	Confidence level	Emotion-based	Exam difficulty	First exam in course	Instructor	Interest in course	Note taking	Outside (non-academic) Factors	Overall Course Performance	Past Performance in other Courses	Performance goal	Preparation behaviors	Previous Exam Performance	Know what to expect	Retention of material	Study guide			Test taking skills
reason exam 2.docx		1	8	25	24	7	8		1	1	7	1	1	6	3	22	62	11	3	2		18	211
reason exam 1.docx	1	3		45	17	4	13	30	1	5	1		2	29	9	52			7	6	5	21	251
Totals	1	4	8	70	41	11	21	30	2	6	8	1	3	35	12	74	62	11	10	8	5	39	

Fig. 3 Frequencies of main reason for performance estimates on Exam 1 and Exam 2

(24 of respondents; sample excerpt: “*I feel confident in getting an A*”). See Fig. 3 for a complete list of frequencies of explanations for performance estimates provided for both exams.

Discussion

As in Study 1, Study 2 results indicated that students tended to overestimate their upcoming exam performance. Whereas this was the case in Study 1 among students in a 200-level course one month prior to the exam, we saw similar levels of overestimation in Study 2 among students from multiple 300-level courses one week before the exam. Specifically, participants overestimated their Exam 1 performance by about 3% overall.

When asked what those performance estimates were based on, results from our qualitative analysis revealed the most common explanations were preparation behaviors, comprehension of the course material, and in the case of Exam 2, prior exam performance in the course. These explanations are somewhat consistent with themes for exam performance predictions found in prior research. For example, comprehension of course material and performance on prior tests were common explanations that emerged for exam predictions among students in Hacker et al.’s (2008) study. Our findings extend these explanatory patterns for performance predictions beyond a single course to multiple courses of varying topics.

Regarding the impact of incentives for prediction accuracy, Study 2 findings revealed that providing a 2-point reward incentive was associated with significantly better accuracy in predicting Exam 2 performance compared to receiving no incentive for accuracy in predictions.

Moreover, when we assessed within-group changes in prediction accuracy from pre-incentive to post-incentive estimates, the 2-point accuracy group showed significant improvement.

Contrary to our expectation, the 5-point reward incentive did not result in greater accuracy than the no incentive condition, nor did the 5-point incentive group show improved accuracy in pre- to post-incentive comparisons. We had anticipated that the 5-point reward incentive group would show similar benefits to the 2-point incentive group. However, it is possible that the 5-point reward incentive seemed “too good to be true” to students and perhaps the students did not believe that they would receive such a large point increase to their exam grade. Conversely, the 2-point incentive was potentially more believable and that may explain why that magnitude of incentive was most effective within this academic context.

In addition to better accuracy than the other groups and significantly improved accuracy in performance predictions, students who received the 2-point incentive for accuracy in performance predictions also experienced a significant improvement in their academic performance. These students improved by about 3% from Exam 1 to Exam 2. As in Study 1, these results provide support for the Dunning-Kruger effect that accurate predictions are associated with better performance. Conversely, students with less accurate predictions experienced poorer performance (Kruger & Dunning, 1999). Dunning et al. (2003) explain that this association may be due to the skill set required to evaluate the accuracy of one’s exam performance may be the same skill set needed to produce correct responses on the exam.

General discussion

The current studies examined personal performance estimates and whether accuracy of those performance predictions could be improved through administering various incentives within an achievement context. Collectively, the findings suggest that initial predictions are unrealistically optimistic, however, performance estimates are malleable and with the appropriate incentives, accuracy in performance predictions can be increased, which in turn, is associated with better actual performance.

Both studies showed initial unrealistic optimism regarding future exam performance prior to any incentive for accuracy. That is, students predicted they would achieve an exam score that was 3% higher than their actual score. Thus, the tendency toward overestimated performance predictions extended beyond one class to both 200 level and various 300 level courses, and was present at one month before the event (in Study 1) and one week before the event (in Study 2) across students from various different academic majors and college levels. These findings are in line with prior research showing unrealistic optimism regarding future events in various domains (see Shepperd et al., 2017 for review) and at various time points prior to the event itself (Ruthig et al., 2017b).

Although this overestimation may motivate certain students to some degree (Zhang & Fishbach, 2010), it may also leave many students unprepared, hence the need to consider means to improve accuracy in performance estimates. Accordingly, we compared the impact of positive reward incentives to negative punishment incentives on accuracy of performance estimates in Study 1, then varied the magnitude of positive reward incentives in Study 2 to identify the optimal amount of incentive to motivate accuracy in performance predictions.

Overall, the results from both studies showed that an accuracy incentive in the form of a punishment incentive in Study 1 and a 2-point reward incentive in Study 2 seemed to be better than no incentive at all in producing more accurate performance predictions. Within Study 2, too much reward incentive seemed to impede the effectiveness of the incentive. Perhaps this is because students did not believe their instructors would actually add five bonus points to their exam scores. Given that, the smaller two-point incentive, though modest, was more believable and yet still important and worth careful consideration of one's upcoming performance. It is possible that this modest incentive was also effective because the estimate concerned a higher stakes situation (i.e., an actual course exam) rather than lower stakes hypothetical situations (e.g., scavenger hunt performance; Amor & Sackett, 2006).

Not only did the modest but meaningful two-point reward incentive significantly improve accuracy and result in better accuracy compared to a bigger incentive, it was also associated with better overall performance. Specifically, students provided with a two-point accuracy incentive outperformed the other groups on their course exam and showed marked improvement in their performance from the first exam to their second exam. This improvement was not demonstrated by the other accuracy incentive groups or by the control/no incentive groups. The possibility of greater accuracy in performance estimates directly contributing to better preparedness and better exam performance should be examined in subsequent research.

Finally, qualitative results from Study 2 showed students commonly based their performance estimates on their preparation behaviors such as attending class, taking notes, and studying, along with their comprehension of the course material, and for the second exam, many students based their estimates on past exam performance in the course. These themes may highlight the need for closer examination of students' preparatory behaviors and comprehension, particularly among students who are not performing as well as they expect. For example, these students may believe they are effective notetakers, but are actually missing key points in their class notes. Similarly, they may perceive themselves as spending a considerable amount of time studying, yet in reality may not be effectively engaging with the material in a meaningful way. Greater examination of these preparatory behaviors and course comprehension are needed. Nonetheless, the current findings provide a glimpse into the basis for performance predictions and directions for further exploration.

Implications

The current studies determined the effectiveness of providing optimal incentives to facilitate accuracy in performance predictions and later performance success in an academic setting. Findings have implications to facilitate instruction and student learning that is conducive to successful academic development in higher education. Specifically, identifying social cognitions such as performance estimates that are associated with students' actual academic performance and in turn, their persistence in college provides useful information that may facilitate a favorable shift in how undergraduate students approach their courses. For example, instructors can provide their students with informational strategies aimed at maximizing a realistic view of their performance which may be associated with students successfully adjusting to the academic demands of college.

Through modifying performance predictions to improve their accuracy and better prepare students to succeed, the

current research also contributes to identifying early indicators of risk (e.g., overestimated performance expectations resulting in under-preparedness and subsequent failure) that can benefit performance and persistence in college. Identifying risk factors early in a college student's academic training is vital to facilitating the student's persistence and degree completion (Singell & Waddell, 2010).

Limitations and future research

Although the current findings provide important insight into the role of accuracy in performance estimates among college students, they are not without some limitations. Specifically, the optimal timing of when to provide an incentive remains unclear. That is, it is unknown whether the effectiveness of an accuracy incentive varies as a function of prediction timing. Subsequent research should address this limitation by directly comparing the same incentive effectiveness a month prior to a performance event vs. a week prior or immediately prior, vs. after the event while awaiting performance feedback.

The present findings are only representative of a small slice of academic performance settings (e.g., undergraduate courses within a single university). Despite multiple majors, college levels, instructors, and courses represented, the extent to which these findings would replicate among larger samples within other academic settings (e.g., humanities courses, medical courses, graduate or professional degrees, etc.), across various higher education institutions, and beyond to occupational or athletic settings needs to be determined. If the findings do extend to other domains, pinpointing the optimal incentive will be an asset to successfully estimating one's upcoming performance and perhaps enhance that performance in various domains.

Author contribution Both authors were involved in all parts of this research.

Funding The authors did not receive support from any organization for the submitted work.

Data availability Data and materials from this study will be made available upon request through a permanent repository at the authors' institution: <https://commons.und.edu/data/>.

Declarations

Treatment of participants in the sample upon which the manuscript is based was in accordance with APA standards. The University of North Dakota's institutional review board approved the project. All participants indicated their consent to participate in writing.

Conflict of interests The authors have no financial or non-financial interests to disclose and have no conflict of interests directly or indirectly related to this work.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Armor, D. A., & Sackett, A. M. (2006). Accuracy, error, and bias in predictions for real versus hypothetical events. *Journal of Personality and Social Psychology*, 91(4), 583–600. <https://doi.org/10.1037/0022-3514.91.4.583>
- Armor, D. A., & Taylor, S. E. (2002). When predictions fail: The dilemma of unrealistic optimism. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (p. 334–347). Cambridge. <https://doi.org/10.1017/CBO9780511808098.021>
- Armor, D. A., & Taylor, S. E. (2003). The effects of mindset on behavior: Self-regulation in deliberative and implemental frames of mind. *Personality and Social Psychology Bulletin*, 29(1), 86–95. <https://doi.org/10.1177/0146167202238374>
- Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, 67(3), 366–381. <https://doi.org/10.1037/0022-3514.67.3.366>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in Psychology. *Psychological Assessment*, 6, 284–290.
- Dedoose Version 9.0.17. (2021). Cloud application for managing, analyzing, and presenting qualitative and mixed method research data. SocioCultural Research Consultants, LLC <https://www.dedoose.com>
- De Vries, H., Elliott, M. N., Kanouse, D. E., & Teleki, S. S. (2008). Using pooled kappa to summarize interrater agreement across many items. *Field Methods*, 20(3), 272–282. <https://doi.org/10.1177/1525822X08317166>
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5(3), 69–106. <https://doi.org/10.1111/j.1529-1006.2004.00018>
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12, 83–86.
- 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 Learning*, 12, 1–19.
- Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition Learning*, 3, 101–121.
- Jansen, R. A., Rafferty, A. N., & Griffiths, T. L. (2021). A rational model of the Dunning Kruger effect supports insensitivity to evidence in low performers. *Nature Human Behaviour*, 5(6), 756–763.
- Jefferson, A., Bortolotti, L., & Kuzmanovic, B. (2017). What is unrealistic optimism? *Consciousness and Cognition*, 50, 3–11.

- 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 Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Lewine, R., & Sommers, A. A. (2016). Unrealistic optimism in the pursuit of academic success. *International Journal for the Scholarship of Teaching and Learning*, 10(2), n2.
- Li, S., Miller, J. E., O'Rourke Stuart, J., Jules, S. J., Scherer, A. M., Smith, A. R., & Windschitl, P. D. (2021). The effects of toll comparisons when estimating the likelihood of task success. *Judgment and Decision Making*, 16, 165–200.
- Ruthig, J. C., Gamblin, B. W., Jones, K., Vanderzanden, K., & Kehn, A. (2017a). Concurrently examining unrealistic absolute and comparative optimism: Temporal shifts, individual difference and event-specific correlates, and behavioural outcomes. *British Journal of Psychology*, 108(1), 107–126. <https://doi.org/10.1111/bjop.12180>
- Ruthig, J. C., Jones, K., Vanderzanden, K., Gamblin, B. W., & Kehn, A. (2017b). Learning from one's mistakes: Understanding changes in performance estimates as a function of experience, evaluative feedback, and post-feedback emotions. *Social Psychology*, 48(4), 185. <https://doi.org/10.1027/1864-9335/a000307>
- Ruthig, J. C., Kroke, A. M., & Holfeld, B. (2022). Anticipating academic performance and feedback: Unrealistically optimistic temporal shifts in performance estimates and primary and secondary control strategies. *Social Psychology of Education*, 25(1), 55–73.
- Shepperd, J. A., Klein, W. M. P., Waters, E. A., & Weinstein, N. D. (2013). Taking stock of unrealistic optimism. *Perspectives on Psychological Science*, 8(4), 395–411. <https://doi.org/10.1177/1745691613485247>
- Shepperd, J. A., Ouellette, J. A., & Fernandez, J. K. (1996). Abandoning unrealistic optimism: Performance estimates and the temporal proximity of self-relevant feedback. *Journal of Personality and Social Psychology*, 70(4), 844–855. <https://doi.org/10.1037/0022-3514.70.4.844>
- Shepperd, J. A., Pogge, G., & Howell, J. L. (2017). Assessing the consequences of unrealistic optimism: Challenges and recommendations. *Consciousness and Cognition*, 50, 69–78.
- Simmons, J. P., & Massey, C. (2012). Is optimism real? *Journal of Experimental Psychology: General*, 141(4), 630–634. <https://doi.org/10.1037/a0027405>
- Simons, D. J. (2013). Unskilled and optimistic: Overconfident predictions despite calibrated knowledge of relative skill. *Psychonomic Bulletin Review*, 20, 601–607.
- Singell, L. D., & Waddell, G. R. (2010). Modeling retention at a large public university: Can at-risk students be identified early enough to treat? *Research in Higher Education*, 51, 546–572.
- Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 39(5), 806–820. <https://doi.org/10.1037/0022-3514.39.5.806>
- Zhang, Y., & Fishbach, A. (2010). Counteracting obstacles with optimistic predictions. *Journal of Experimental Psychology: General*, 139(1), 16–31. <https://doi.org/10.1037/a0018143>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.