

# Unskilled and unaware in the classroom: College students' desired grades predict their biased grade predictions

Michael J. Serra<sup>1</sup> · Kenneth G. DeMarree<sup>2</sup>

© Psychonomic Society, Inc. 2016

**Abstract** People tend to be overconfident when predicting their performance on a variety of physical and mental tasks (i.e., they predict they will perform better than they actually do). Such a pattern is commonly found in educational settings, in which many students greatly overestimate how well they will perform on exams. In particular, the lowest-performing students tend to show the greatest overconfidence (i.e., the “unskilled-and-unaware” effect). Such overconfidence can have deleterious effects on the efficacy of students’ short-term study behaviors (i.e., underpreparing for exams) and long-term academic decisions (i.e., changing one’s academic major to an “easier” topic or dropping out of school completely). To help understand why students’ grade predictions are often overconfident, we examined the hypothesis that students’ grade predictions are biased by their desired levels of performance, which are often much higher than their actual levels of performance. Across three studies in which actual students made predictions about their exam performance in their courses, we demonstrated that students’ grade predictions are highly biased by their desired grades on those exams. We obtained this result when students predicted their exam grades over a week before the exam (Study 1), immediately after taking the exam (Study 2), and across the four course exams in a single semester (Study 3). These results are informative for understanding why the “unskilled-and-unaware” pattern of performance predictions occurs, and why people

in general tend to be overconfident when making both physical and mental performance predictions.

**Keywords** Metacognition · Overconfidence · Unskilled and unaware · Wishful thinking · Predictions of performance

## Introduction

People tend to be overconfident in their own abilities and in their future performance on a variety of tasks. For example, adults overestimate their skills and performance on tasks related to reasoning, humor, and grammar (Kruger & Dunning, 1999); children overestimate their ability to remember pictures (Lipko, Dunlosky, & Merriman, 2009) and to perform physical tasks (Schneider, 1998); and consumers overestimate how easy it will be to learn to use a new product (Billeter, Kalra, & Loewenstein, 2011). Most relevant to the present research, students often overestimate how well they will perform on an upcoming test of their learning (e.g., Hacker, Bol, & Bahbahani, 2008; Miller & Geraci, 2011a, b; but see Griffin, Jee, & Wiley, 2009; Shanks & Serra, 2014), which can impair their study behaviors (Metcalfe & Finn, 2008). The reasons for such overconfidence, however, are not yet well understood. Toward this end, in the present studies we tested the hypothesis that students make overconfident predictions of their test performance because their predictions are related to—and potentially biased by—their *desired* level of performance, which is typically higher than their actual level of performance.

---

✉ Michael J. Serra  
michael.serra@ttu.edu

<sup>1</sup> Department of Psychological Sciences, Texas Tech University, MS 2051, Psychology Building, Lubbock, TX 79409-2051, USA

<sup>2</sup> Department of Psychology, University at Buffalo, State University of New York, Buffalo, NY, USA

## Persistent overconfidence

Researchers in a variety of domains have examined people’s self-assessments of their own abilities and predictions of their own future performance for both mental and physical tasks. In

the context of human learning within formal-education settings, people's thoughts about their thoughts (i.e., *metacognition*; see Briñol & DeMarree, 2012; Dunlosky & Metcalfe, 2009) play a major role in determining the efficacy of students' study behaviors. For example, students utilize *metacognitive knowledge* when selecting strategies for how to study, engage in *metacognitive monitoring* when evaluating how well they already know what they are studying, and engage *metacognitive control* when using their knowledge and monitoring to decide for how long to continue to study, or whether to stop studying altogether (i.e., Nelson & Narens, 1990; for a review, see Dunlosky & Metcalfe, 2009). Critically, flaws in any of these components of metacognition can impair students' study behaviors (Serra & Metcalfe, 2009).

In the present studies, we examined a common error in students' monitoring of their learning and predictions of their future test performance: the persistent *overconfidence* of students' judgments (e.g., Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008; Erev, Wallsten, & Budescu, 1994; Klayman, Soll, González-Vallejo, & Barlas, 1999; Koriat, Lichtenstein, & Fischhoff, 1980). Overconfidence in this regard (i.e., students' predictions of their future test performance being significantly higher than their actual performance) can be problematic for students, because it can lead them to underprepare for exams (e.g., Bjork, Dunlosky, & Kornell, 2013; Dunlosky & Rawson, 2012; Grimes, 2002; Kornell & Bjork, 2008; Kornell & Metcalfe, 2006; but see Carpenter, Wilford, Kornell, & Mullaney, 2013) or make poor restudy decisions (cf. Finn, 2008; Metcalfe & Finn, 2008; Shanks & Serra, 2014).

### Unskilled and unaware

Although it could prove problematic for *any* student to overestimate their learning (especially leading up to an exam), of further concern is the common observation that the lowest-performing (and, perhaps, lowest-ability) students are the ones most overconfident in their judgments, whereas the highest-performing students tend to be more accurate, or even demonstrate underconfidence (for a review, see Miller & Geraci, 2011b). This pattern is often referred to as “unskilled and unaware,” because it seems that the lowest-performing people in a domain not only lack the abilities or knowledge to perform well in that domain, but also lack metacognitive awareness of their deficits (e.g., Dunning, Johnson, Ehrlinger, & Kruger, 2003; Grimes, 2002; Kruger & Dunning, 1999; but see Griffin et al., 2009; Miller & Geraci, 2011b; Shanks & Serra, 2014). In formal-education settings, this pattern can be especially problematic. This pattern not only might lead the lowest-performing students to underprepare for exams within their courses, but also could suggest to these students that they

are not capable of passing their courses, that they cannot achieve the level of performance they expected in those or other courses, and even that they are not at all capable of succeeding in that field of study (cf. Grimes, 2002). For example, Stinebrickner and Stinebrickner (2014) reported that students tend to change their major from a science to a nonscience field because their grades in their science courses were substantially lower than they initially expected.

To date, a number of researchers have attempted to uncover the factor(s) that produce the unskilled-and-unaware pattern (e.g., Ackerman, Beier, & Bowen, 2002; Burson, Larrick, & Klayman, 2006; Dunning et al., 2003; Ehrlinger et al., 2008; Kruger & Dunning, 2002; Krueger & Mueller, 2002; see also Miller & Geraci, 2011a, b). Although a variety of factors have been identified as contributing to this pattern (and some have been eliminated from contention), no one factor has been identified as being the major, or sole, cause of this outcome. In the present studies, we examined another potential contributing factor: students' *desired* grades. Kruger and Dunning (1999) noted that motivational biases likely play a role in generating the unskilled-and-unaware effect (p. 1132), but no research that we are aware of has empirically examined this idea. To address this empirical deficiency, in the present studies we examined students' desired exam grades and assessed whether those desired grades were associated with students' exam grade predictions in the classroom.

### Wishful thinking

The idea that students' desired grades might impact their predictions of the grades they might actually earn—and cause these predictions to be overconfident—is supported by the empirical literature on *wishful thinking*. According to this account, people unintentionally base their predictions of future performance on how they *want* to perform rather than on how they *expect* to perform (e.g., Schneider, 1998; Stipek, Roberts, & Sanborn, 1984; Willard & Gramzow, 2009). For example, young children (i.e., preschoolers and kindergartners) continue to overestimate their future performance even when they have had experience with the target task and can remember their past performance accurately (e.g., Lipko, Dunlosky, & Merriman, 2009; Schneider, 1998). This persistent overconfidence seems to stem from wishful thinking rather than a general inability to factor past performance into predictions of future performance, since young children often do not demonstrate overconfidence when asked to make predictions of future performance for another child whose past performance they are aware of (e.g., Schneider, 1998; but see Lipko et al., 2009). Put differently, young children make better predictions of *other* children's performance than of their own because their own predictions are biased by their own goals, which typically involve a desire for stellar performance.

Although college students' grade predictions are not always accurate, in general adults do not show the very large and persistent effects of wishful thinking that children demonstrate (e.g., college students heavily factor past-test performance into their current learning predictions; see Ariel & Dunlosky, 2011; England & Serra, 2012; Finn & Metcalfe, 2007; Tauber & Rhodes, 2012). Nevertheless, we still propose that wishful thinking might exert an effect on adults' performance predictions in a more subtle way, and there is some support for this prediction. Soderstrom and McCabe (2011) had college students study and then make judgments of learning (JOLs) for paired associates before taking a memory test over those items. Participants received between one and six points for recalling each item (point values were randomly assigned to the items). Most importantly, in one condition participants were told the number of points that each item would be worth *in between* the study and judgment of each item. Although the point value did not affect participants' recall in this condition (because study had occurred before the point value was known), participants judged that their memory would be better for items worth more points than for items worth fewer points. Given that the participants in this condition could do nothing to influence their likelihood of recall at that point in time, it seems likely that their *desire* to earn more points led them to judge higher-point items to be more memorable than lower-point items—a wishful-thinking effect.

Soderstrom and McCabe (2011) offered compelling initial support for our prediction, but the context in which they examined the effects of wishful thinking was quite different from the one we proposed to examine, so there was no guarantee that desires would operate the same way in both situations. More specifically, whereas they examined item-by-item JOLs (made after studying each of several dozen items), our focus in the present studies was mostly on predictions of overall exam performance (single global performance predictions). It was not clear whether the effects obtained with item-by-item memory predictions would also occur for global predictions of performance like those in the present studies; item-by-item predictions might be less prone to the biasing effects of desires than global predictions because the former are based heavily on the item-specific properties of those items (e.g., difficulty, see Koriat, 1997). Furthermore, we examined students' predictions of performance that they made either before (Study 1) or after (Studies 2 and 3) the exams had occurred; it is not known whether the effects obtained with one of these types of prediction would also occur for the other type. For example, people's metacognitive experience while learning or studying provides the major basis for performance predictions made before a test, whereas people's metacognitive experience while taking the exam (i.e., during recall) provides the major basis for

postexam performance predictions. Pretest judgments might be more susceptible to the effects of desires than posttest judgments are because people could base the latter type of judgment on their actual test experience. In addition, the "real-life" nature of the predictions that participants make for actual exams might involve greater thoughtfulness than predictions they make for a no-consequence laboratory exercise, and motivations might have greater relevance for "real-life" than for artificial predictions. Either of these possibilities might actually *increase* the likelihood of obtaining a relationship between desires and predictions for actual exams as compared to laboratory tasks.

### The present studies

The purpose of the present studies was to examine the hypothesis that students' exam predictions are typically overconfident because their predictions are heavily biased by their *desired* grade on the exam, which is fairly high. This explanation might be especially important for helping identify the factors that contribute to the unskilled-and-unaware pattern of performance predictions. To examine these issues, we had students enrolled in either a general psychology course (Study 1) or a research methods course (Studies 2 and 3) report their *desired* grades in addition to reporting the grades they actually *expected* to earn. If desired grades lead to inflated predictions of performance, then students' predictions of their grades should be related not only to their actual performance, but also to their desired performance. Note that in each of our studies, students were asked to predict their absolute grades rather than their relative performance, to reduce the influence of people's assessments of others' performance on the unskilled-and-unaware phenomenon (Hartwig & Dunlosky, 2014).

### Study 1

The purpose of Study 1 was to obtain some initial evidence to support our hypothesis that students' desired grades lead to inflated predictions of their actual grades. To do so, we had students report their desired and expected grades for both the final exam and the overall course. The latter judgments would provide a potentially more conservative test of our hypothesis than the former, because students can presumably use a wealth of information acquired over the course of the semester to predict their final course grade. This study also allowed us to compare results involving students' actual grade predictions in a classroom setting to similar studies obtained in the laboratory utilizing item-by-item judgments (i.e., Soderstrom & McCabe, 2011).

## Method

**Participants** The participants in the present sample were 485 college students enrolled in General Psychology at Texas Tech University, who participated for course credit. In this and all of these studies, the sample size was determined by the number of students enrolled in the course who were willing to provide data for the present purposes. More specifically, the sample was determined by asking all students in a given class to provide the information relevant to the present hypotheses. Although not all students chose to provide these estimates and desires, the vast majority of the students in each class did provide the information. This sample was 60% female and 40% male, with a modal age of 19 years old. Furthermore, including people who indicated more than one race or ethnic group, the sample was 73.4% White or European American, 20.6% Hispanic or Latino, 6.2% Black or African American, 3.5% Asian or Asian American, 1.6% Native American or Alaskan Native, and 2.1% some other race or group.

**Procedure** Students completed the present questions as part of a larger end-of-semester course assessment. (Other measures related to course outcomes were collected at this time but were not directly relevant to the present purposes, and hence are not reported.) We collected all data online using the Qualtrics survey software during the last two weeks of the semester, before final exams were administered but after most other assignments and preceding exams in the course had been completed.

During the critical section of the assessment, students answered two questions about their expected performance on the final exam for the course: “What grade do you realistically *want* to earn on your final exam in this course?” (*desired grade*) and “What grade do you think you will *actually* earn on your final exam in this course?” (*predicted grade*). We included the word “realistically” in the phrasing of our desired-grade questions to prevent respondents from passively answering “100%.” We wanted them to indicate their desires, but to make predictions within the realm of their own perceived possibilities. For these questions, students were instructed to enter a whole number between 0 and 100 to indicate their expected grade out of 100% (and the survey software only accepted such values). Students then answered the same two questions, restated to be about their final grade in the course as a whole.

Note that although the exact format of the final exams (as well as the exams and assignments utilized during the semester that contributed to the final course grades) varied by instructor, for the most part there was high consistency across the sections of the course. In particular, the final exams were typically noncumulative and consisted primarily of multiple-choice questions.

Students were made aware that their responses would later be compared to their actual grades and were allowed to skip any of the questions if they chose to. Students also provided us with permission (or denied us permission) to access their final exam and course grades for the purpose of comparing them to their predictions. As such, the number of students contributing data to some of the analyses below varies slightly from measure to measure. In general, however, most students completed all of the present measures. An additional 26 students completed at least some of the course assessment but did not complete any of the present measures, and as such were excluded from the present sample (since they had no data to contribute). We then obtained final exam and course grades from the course instructors for comparison to the students’ predictions for those students who had completed at least some of the present measures and who provided us with permission to access these grades. After that point, the data were completely deidentified for analysis.

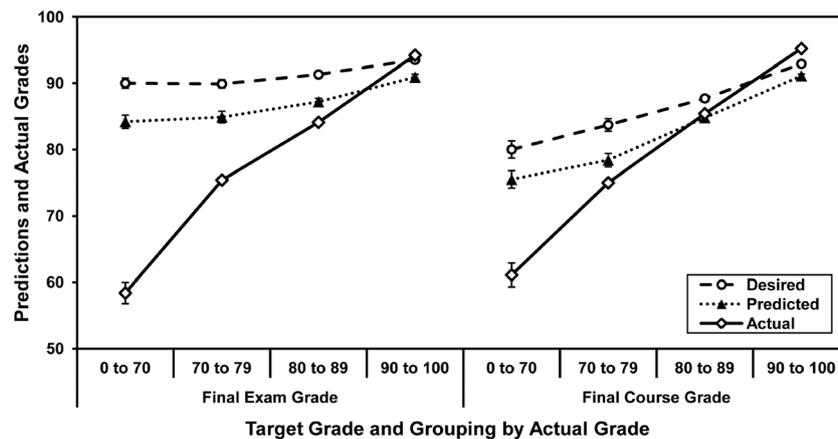
## Results and discussion

For descriptive information only, we separated students into two sets of four groups based on (a) their actual scores on the final exam and (b) their actual final course grades, much as in Miller and Geraci (2011b). Both Table 1 and Fig. 1 present the means of students’ desired, predicted, and actual grades. Although both Table 1 and Fig. 1 present grouped data, all subsequent analyses used regression with continuous predictors and outcomes, unless otherwise reported. The values in the table and figure are provided mainly to illustrate the general pattern of results and for direct comparison to the results

**Table 1** Descriptive statistics by group in Study 1

| Group              | <i>n</i> | Desired  |               | Predicted |               | Actual   |               |
|--------------------|----------|----------|---------------|-----------|---------------|----------|---------------|
|                    |          | <i>M</i> | ( <i>SD</i> ) | <i>M</i>  | ( <i>SD</i> ) | <i>M</i> | ( <i>SD</i> ) |
| Final Exam Grade   |          |          |               |           |               |          |               |
| 0–70               | 71       | 90.0     | (6.4)         | 84.2      | (8.4)         | 58.4     | (13.7)        |
| 70–79              | 114      | 89.9     | (6.6)         | 84.9      | (9.4)         | 75.4     | (2.9)         |
| 80–89              | 168      | 91.3     | (5.8)         | 87.2      | (6.6)         | 84.1     | (2.9)         |
| 90–100             | 132      | 93.5     | (4.4)         | 90.9      | (5.1)         | 94.2     | (3.5)         |
| Final Course Grade |          |          |               |           |               |          |               |
| 0–70               | 29       | 80.0     | (6.9)         | 75.5      | (6.7)         | 61.1     | (9.6)         |
| 70–79              | 66       | 83.7     | (7.7)         | 78.4      | (8.4)         | 75.0     | (2.9)         |
| 80–89              | 131      | 87.7     | (5.8)         | 84.8      | (6.2)         | 85.4     | (2.8)         |
| 90–100             | 259      | 92.9     | (3.8)         | 91.1      | (4.5)         | 95.2     | (3.5)         |

We grouped students by their scores on the final exam and separately by their final course grade. These groupings are shown only for descriptive purposes and to aid comparison to other, related studies; all critical analyses (i.e., regression analyses) involved the analysis of continuous data. The means represent students’ desired, predicted, and actual grades on the final exam and for the course



**Fig. 1** Mean desired, predicted, and actual final exam and final course grades by target-grade (final exam vs. course grade) and actual-grade grouping in Study 1. Error bars show the standard errors of the means (note that many are too small to be seen at this scale). These groupings are

of other, similar studies (e.g., Miller & Geraci, 2011b). We also chose to group students on the basis of their actual grades rather than quartiles because grouping by quartile is inappropriate if one is to consider absolute accuracy or calibration at the level of actual grades. Although we grouped by grade for descriptive purposes (i.e., to more easily show that the unskilled-and-unaware pattern was obtained in the present studies), we conducted all analyses utilizing regression with continuous predictors.

**Final exam grade** To examine students' potential projection of desires onto their grade predictions on the final exam, we simultaneously regressed the exam predictions on desired exam grade and actual exam grade. Students' actual grades were related to their predicted grades ( $b = .113$ ,  $se = .021$ ,  $\beta = .190$ ),  $t(477) = 5.32$ ,  $p < .001$ , but their desired grade was a strong, significant predictor as well ( $b = .758$ ,  $se = .047$ ,  $\beta = .581$ ),  $t(477) = 16.27$ ,  $p < .001$ . Thus, both actual performance and desired performance were independent, significant predictors of grade predictions. This observation is anecdotally supported by the grouped data plotted in Fig. 1 (left panel), in which desired grades track grade predictions more closely than do actual grades. These results—which demonstrate a strong relationship between students' desired and predicted grades—support our hypothesis that students' desired grades lead to inflated predictions of their exam performance.<sup>1</sup>

<sup>1</sup> It is possible that asking respondents to provide their desired grade before their predicted grade inflated the present effect sizes by making the potential anchor (i.e., desired grades) especially salient (i.e., Study 1). In our Studies 2 and 3, however, respondents made their grade predictions 30–90 min after they had indicated their desired grade, with the actual exam intervening between the two time points at which they made these two judgments. It does not seem likely that any effect of this ordering would be maintained over this filled delay.

shown only for descriptive purposes and to aid comparison to other, related studies; all critical analyses (i.e., regression analyses) involved the analysis of continuous data

**Final course grade** To examine students' potential projection of desires onto their final course-grade predictions, we simultaneously regressed the course-grade predictions on desired course grade and actual course grade. Students' actual grades were related to their predicted grades ( $b = .223$ ,  $se = .023$ ,  $\beta = .299$ ),  $t(482) = 9.58$ ,  $p < .001$ , but—as before—their desired grades were also strong predictors ( $b = .728$ ,  $se = .037$ ,  $\beta = .621$ ),  $t(482) = 19.88$ ,  $p < .001$  (see Fig. 1, right panel). Again, desired grades tracked grade predictions more closely than did actual grades. As with the exam grades, these results support our hypothesis that students' desired grades lead to inflated predictions of their final course performance.

The results pertaining to final course grades are, perhaps, more valuable than those pertaining to final exam grades, because students had a larger variety of information that they could have used to make their final course-grade predictions than to make their final exam predictions. Specifically, although past-test performance in the course *might* have provided some informative cues on which to base final exam predictions, that generalization would only be useful if students assumed that all exams in the course were of equal difficulty. In contrast, students' final course grades were determined by the direct combination of their grades on exams, quizzes, and assignments throughout the semester; at the point at which we collected the present data, students should have been more or less aware of the final grade they'd likely earn in the course on the basis of their grades thus far. Nevertheless, students' predictions of their course grades demonstrated the unskilled-and-unaware pattern and were highly related to—and therefore potentially biased by—their desired grades (see also Soderstrom & McCabe, 2011).

## Study 2

Although Study 1 provided initial support for our predictions, it has a critical weakness. Specifically, because students made their predictions *before* they took the exam, their predictions could not be impacted by their experience in taking the exam, and may instead have been affected by overly optimistic plans to study that had been impacted by desired grades, leading to a potentially inflated impact of desired grades. In Study 2, we included the present measures in conjunction with an actual course examination. In this case, we had students indicate their *desired* grade just prior to taking the exam, but had them indicate their *predicted* grade immediately after completing the exam. Students' postdictions of test performance are generally more accurate than their predictions (e.g., Glenberg & Epstein, 1987; Maki, 1998; Winne & Jamieson-Noel, 2002); thus, this study would offer a stronger test of the role that desired grades might play in students' performance estimates. One possibility is that the subjective experience of taking the exam is particularly salient when students make their postexam predictions, and this could increase their ability to predict actual test performance free of the biasing influence of their desired grades. In contrast, another possibility is that students' desires bias their interpretation of the metacognitive cues they use to make predictions of their performance, including their very recent experience of taking the exam (cf. Balcetis & Dunning, 2006, 2010; Dunning & Balcetis, 2013). If so, students' grade predictions would continue to be biased by their desires.

## Method

**Participants** The participants in the present sample were 66 college students enrolled in a research methods course at Texas Tech University. They completed these measures as ancillary questions on their final exam. We did not collect demographic information for these students, since this analysis was not part of a larger course assessment, but the sample characteristics were likely similar to those in Study 1. Seven of the 66 students in the course did not complete all of the present measures, so the effective sample size was 59.

**Procedure** Students completed the present questions as ancillary measures on their final examination. The final exam was noncumulative, covering the final quarter of a methods course. The exam consisted of 44 multiple-choice questions and two short-answer questions. Before students began the exam, they answered the question, "What percent grade (0–100%) do you (realistically) *want* to earn on this exam?" (desired grade). After completing the exam, students answered the question, "Now that you have taken this exam, what percent grade (0–100%) do you think you *actually* earned on this exam (not

including extra credit)?" (predicted grade). For these questions, students were instructed to provide a whole number between 0 and 100. Students were made aware at the start of the exam that their responses would later be compared to their actual grades, and they were allowed to skip any of the questions if they chose to. In general, however, most students (59 of 66) completed all of the present measures. Once the exams had been graded, the numbers described above, along with students' actual exam scores, were entered into a deidentified data file for analysis.

## Results and discussion

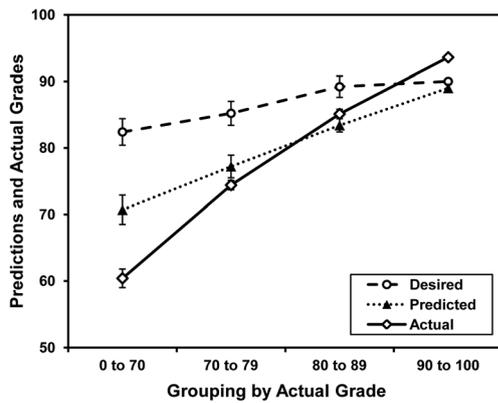
For descriptive information, we separated students into four groups based on their actual scores on the final exam, as in Study 1 and previous studies (i.e., Miller & Geraci, 2011b). Both Table 2 and Fig. 2 present the means of students' desired, predicted, and actual grades. Although both Table 2 and Fig. 2 present grouped data, all analyses reported below were based on regression unless otherwise reported.

To examine students' potential projection of desires onto grade predictions on the final exam, we regressed their exam predictions on desired exam grade and actual exam grade. Given that we had the students in Study 2 make their grade predictions *after* completing the exam, it is perhaps not surprising that their actual grades related to their predicted grades ( $b = .358$ ,  $se = .085$ ,  $\beta = .420$ ),  $t(56) = 4.22$ ,  $p < .001$ . Importantly, however, students' desired grades were also related to their predicted grades ( $b = .532$ ,  $se = .117$ ,  $\beta = .451$ ),  $t(56) = 4.53$ ,  $p < .001$ . Thus, although students had just completed taking the exam—and thus their subjective experience on the exam should have been particularly salient—their predicted exam scores were still strongly related to their desired grades. These results provide further support for our hypothesis that students' desired grades lead to inflated predictions of their exam performance.

**Table 2** Descriptive statistics by group in Study 2

| Group  | <i>n</i> | Desired  |               | Predicted |               | Actual   |               |
|--------|----------|----------|---------------|-----------|---------------|----------|---------------|
|        |          | <i>M</i> | ( <i>SD</i> ) | <i>M</i>  | ( <i>SD</i> ) | <i>M</i> | ( <i>SD</i> ) |
| 0–70   | 18       | 82.4     | (8.3)         | 70.7      | (10.3)        | 60.4     | (6.6)         |
| 70–79  | 20       | 85.2     | (8.0)         | 77.2      | (7.7)         | 74.4     | (3.4)         |
| 80–89  | 21       | 89.2     | (7.5)         | 83.4      | (4.6)         | 85.1     | (3.3)         |
| 90–100 | 1        | 90.0     | –             | 89.0      | –             | 93.6     | –             |

We grouped students by their scores on the final exam. These groupings are shown only for descriptive purposes and to aid comparison to other, related related studies; all critical analyses (i.e., regression analyses) involved the analysis of continuous data. The means represent students' desired, predicted, and actual grades on the final exam. Note that only one student scored a 90 or above



**Fig. 2** Mean desired, predicted, and actual final exam grades by actual-grade grouping in Study 2. Error bars show the standard errors of the means (note that only one student scored a 90 or higher, and hence those means have no error bars). These groupings are shown only for descriptive purposes and to aid comparison to other, related studies; all critical analyses (i.e., regression analyses) involved the analysis of continuous data

**Study 3**

In our final study, we sought to replicate and extend the results of our previous studies by combining the strengths of the two studies into a single study. To do this, we had participants report their desired and predicted grades on each of four exams during a semester-long course. Because participants predicted their actual exam grades immediately following each exam (as in Study 2), recent experiential information was available to inform their predictions. But because participants made judgments over the span of a full semester, as the semester progressed they could accumulate knowledge of their performance in the course and on past exams on which to base their estimates (conceptually similar to Study 1). This allowed us to examine the extent to which the influence of people’s desires on their performance estimates persists as their opportunities for knowledge of their actual abilities increased.

**Method**

**Participants** The participants in the present sample were 60 college students enrolled in a research methods course at Texas Tech University, who completed these measures as ancillary questions that appeared on all four exams given in the course during the semester. We did not collect demographic information for these students, since this analysis was not part of a larger course assessment, but the sample characteristics were likely similar to those in Study 1. The exams were non-cumulative, carried equal weights in students’ grades, and were similar in structure to the one described in Study 2. Students were made aware at the start of the exam that their responses would later be compared to their actual grades, and they were allowed to skip any of the questions if they chose to.

Because the lowest exam in this course was dropped from students’ final grades, many students ( $n = 40$ ) opted not to take the fourth exam (i.e., they were satisfied with their grade to date and opted to drop a 0). (Note that the same results would be obtained if the fourth exam were completely excluded from the analyses.) Furthermore, not all students completed all measures each time, leaving observations for 171 occasions (see Table 3 for the statistics on each exam).

**Procedure** The procedure was identical to that of Study 2 (i.e., students reported their desired grade at the start of each exam and their predicted grade at the completion of each exam), except that we collected data for all four exams in the course instead of only for the final exam.

**Results and discussion**

For comparison to the descriptive information from the earlier studies, we first separated students’ predictions into four groups based on their actual scores on each of the four exams. More specifically, Table 4 and Fig. 3 present the mean desired, predicted, and actual exam grades for any student scoring within a given range on any of the four exams in the course (i.e., most students contributed four sets of desired, predicted, and actual exam grades to these means). Although both Table 4 and Fig. 3 present grouped data, all analyses reported below were based on multilevel modeling across participants unless otherwise reported.

Because we had multiple observations for each person (and thus the observations were not independent) and the number of observations per student varied, we employed multilevel modeling for these analyses. In this context, multilevel modeling has the advantages both of a repeated measures analysis of variance, in that it explicitly takes into account the nonindependence of the observations from a given unit (in this case, each student), and of regression, in that it allows for

**Table 3** Descriptive statistics by exam in Study 3

| Exam | <i>n</i> | Desired  |               | Predicted |               | Actual   |               |
|------|----------|----------|---------------|-----------|---------------|----------|---------------|
|      |          | <i>M</i> | ( <i>SD</i> ) | <i>M</i>  | ( <i>SD</i> ) | <i>M</i> | ( <i>SD</i> ) |
| 1    | 55       | 88.8     | (6.6)         | 80.4      | (9.0)         | 79.3     | (11.7)        |
| 2    | 39       | 84.9     | (7.9)         | 75.0      | (11.2)        | 72.4     | (11.9)        |
| 3    | 58       | 84.9     | (7.2)         | 78.3      | (8.8)         | 80.0     | (8.8)         |
| 4    | 19       | 81.7     | (11.7)        | 76.8      | (9.4)         | 70.5     | (9.0)         |

The means represent students’ desired, predicted, and actual grades on a given exam. These groupings are shown only for descriptive purposes and to aid comparison to other, related studies; all critical analyses (i.e., regression analyses) involved the analysis of continuous data. All means represent the values from students who provided complete data, which varied by exam. Note, also, that many students opted not to take the last exam

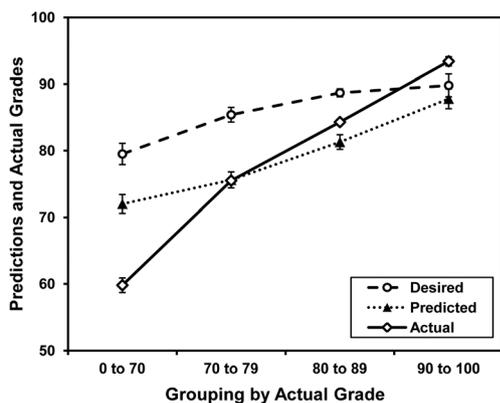
**Table 4** Descriptive statistics by group in Study 3

| Group  | n  | Desired |       | Predicted |       | Actual |       |
|--------|----|---------|-------|-----------|-------|--------|-------|
|        |    | M       | (SD)  | M         | (SD)  | M      | (SD)  |
| 0–70   | 38 | 79.5    | (9.6) | 72.0      | (8.6) | 59.8   | (6.7) |
| 70–79  | 49 | 85.4    | (7.6) | 75.6      | (8.4) | 75.5   | (3.1) |
| 80–89  | 72 | 88.7    | (5.4) | 81.3      | (9.1) | 84.3   | (2.8) |
| 90–100 | 12 | 89.8    | (6.0) | 87.8      | (5.3) | 93.4   | (2.3) |

We grouped students by their scores on a given exam (note that each student contributed up to four sets of exam grades and predictions). These groupings are shown only for descriptive purposes and to aid comparison to other, related studies; all critical analyses (i.e., regression analyses) involved the analysis of continuous data. The means represent students' desired, predicted, and actual grades on a given exam

continuous predictors and provides parameter estimates that are analogous to regression coefficients. We used restricted maximum likelihood estimation and allowed the intercepts—but not the slopes—to vary across analyses (i.e., only the intercept was treated as a random effect). Analyses in which slopes were allowed to vary produced the same patterns of significance.

To examine participants' projection of desires onto their grade predictions on each exam, we first conducted analyses parallel to those we had conducted in Studies 1 and 2, in which students' predicted exam grades had been statistically predicted by their desired and actual exam grades using multilevel modeling. As in Studies 1 and 2, both students' actual grades ( $b = .212$ ,  $se = .054$ ,  $\beta = .216$ ),  $t(166) = 3.97$ ,  $p < .001$ , and their desired grades ( $b = .543$ ,  $se = .069$ ,  $\beta = .389$ ),  $t(152) = 7.86$ ,  $p < .001$ , significantly predicted their predicted grades. Furthermore, as in Study 2, although students had just completed taking the exams, and thus their subjective experience on the exams should have been particularly salient, their



**Fig. 3** Mean desired, predicted, and actual exam grades by actual-grade grouping in Study 3. Note that each student contributed up to four sets of exam grades and predictions. Error bars show the standard errors of the means. These groupings are shown only for descriptive purposes and to aid comparison to other, related studies; all critical analyses (i.e., regression analyses) involved the analysis of continuous data

predicted exam scores were still strongly related to their desired grades. Once again, the results support our hypothesis that students' desired grades lead to inflated predictions of their exam performance.

We also examined whether the impacts of students' actual and desired exam grades varied in their predictive utility across the semester. To do this, we added exam number (coded 1–4) and its interactions with actual and desired grades as linear predictors of predicted grade. This analysis did not alter the results reported above, since exam did not interact with either actual or desired grade, although both trends were such that the effects decreased as the semester progressed ( $t_s < 1.4$ ,  $p_s > .18$ ). Although students' grade predictions did not decrease across the semester, their desired grades did ( $b = -1.98$ ),  $t(124) = 3.99$ ,  $p < .001$ . This finding suggests that students might have adjusted their desires to be within the realm of possibility across the semester as they learned more about their aptitude for the course and the level of difficulty of the exams. Therefore, the present data support the idea that people's predictions of their grades are strongly related to—and therefore possibly biased by—their desired grades. These data do not, however, provide support for the notion that participants learn from their earlier mispredictions and make more accurate predictions as the semester progresses (although it is possible that this form of learning is relatively small, and that a larger sample size could detect such an effect).

## General discussion

People's predictions of their physical and mental performance are often overconfident (e.g., Billeter et al., 2011; Kruger & Dunning, 1999; Lipko et al., 2009; Schneider, 1998). Furthermore, the phenomenon known as “unskilled and unaware” is characterized by a pattern of performance predictions in which the lowest-performing people are the most overconfident in their performance predictions, but the highest-performing people tend to be accurate or to demonstrate underconfidence (e.g., Dunning et al., 2003; Grimes, 2002; Kruger & Dunning, 1999). In three studies, we examined the hypothesis that students' predictions of their grades are persistently overconfident because their predictions are biased by their *desired* level of performance. The results of all three studies strongly supported this hypothesis: Students' desired grades were strong predictors of their grade predictions, when students made their predictions either more than a week before the exam (Study 1) or immediately after completing the exam (Studies 2 and 3).

Overconfidence can have detrimental effects on students' performance on an upcoming exam if they overestimate their learning and, in turn, underprepare for the exam (e.g., Bjork et al., 2013; Dunlosky & Rawson, 2012; Finn, 2008; Grimes, 2002; Kornell & Bjork, 2008; Kornell & Metcalfe, 2006;

Metcalf & Finn, 2008; Shanks & Serra, 2014). Furthermore, in the case of long-term or more general overconfidence (i.e., overconfidence on things in the longer-term future), students' desires might also affect their intended plan of study. Students might anticipate that they will have ample time to study for an exam to earn a high grade on it, but end up overestimating how much time they will actually be willing or able to devote to study as the exam approaches (i.e., *the planning fallacy*; see Buehler, Griffin, & Ross, 2002). Future attempts to make students' JOLs more accurate might address the potential role of desired grades examined in the present studies. We are not suggesting that researchers attempt to reduce the magnitude of students' desired grades; rather, it may be worth exploring whether making students aware that their desires may bias their judgments could reduce these biases (i.e., keep students' desired grades high but reduce students' reliance on these desires when assessing the current state of their learning).

### Might the unskilled be aware?

Despite the frequent occurrence of the unskilled-and-unaware pattern—and concern about its implications for the unskilled—evidence suggests that the lowest-performing people in a given domain have some realization that their estimates might not be accurate. Specifically, Miller and Geraci (2011b) had students rate how confident they were in the accuracy of their exam grade predictions. Despite obtaining the unskilled-and-unaware pattern of results, in which the lowest-performing students made highly overconfident grade predictions, these students were less confident in the accuracy of their predictions than were the highest-performing students. Why, then, did these students make high predictions that they were unsure about?

One possibility is based on the finding that discrepancies between a person's desires and the actual experience while taking the exam can create a sense of conflict, ambivalence, confusion, or lack of clarity. For example, discrepancies between people's actual and desired evaluations lead people to feel more ambivalent in that evaluation (DeMarree, Wheeler, Briñol, & Petty, 2014), and discrepancies between actual and desired levels of self-esteem undermine a person's self-clarity (DeMarree & Rios, 2014). This presumably occurs because, when a person's desired and actual evaluations are discrepant, their choices of behavioral responses to those two evaluations are likely to be in conflict. For example, when people want to be more positive or more negative than they currently are in their evaluation of a political candidate, their actual and desired evaluations might create conflicting approach/avoidance tendencies (DeMarree et al., 2014). In the present context, if people have a high desired grade for an exam but their experience taking the exam leads them to believe they have in fact performed poorly on it, they should hold their high grade prediction with less certainty than if they had desired a more

moderate grade or experienced the exam as being easy. Put differently, the lowest-performing students may often experience a discrepancy between their goal grades and their experience with the exam, whereas the highest-performing students will not. This could lead the lowest-performing students to be relatively less confident in their predictions than the highest-performing students (cf. Händel & Fritzsche, 2016; Miller & Geraci, 2011b).

Another possibility is that students' desired grades might serve as a psychological *anchor*, or mental starting point, for their grade predictions (e.g., England & Serra, 2012; Scheck, Meeter, & Nelson, 2004). After making an initial judgment (based largely on their desired grade), the lowest-performing students might do little to further evaluate or adjust their initial predictions. In contrast, the highest-performing students might also start with a desire-based anchor, but because of their knowledge of that domain, they might then be able to engage in further evaluation, deliberation, and adjustment of their predictions before arriving at a final value. Concordantly, the lowest-performing students should be relatively less confident in their predictions than the highest-performing students (cf. Händel & Fritzsche 2016; Miller & Geraci, 2011b), since different levels of deliberation went into their final evaluations (cf. Barden & Petty, 2008; Dunlosky, Serra, Matvey, & Rawson, 2005; Serra & England, 2012). Judgments made via more thoughtful anchoring and adjusting also tend to have greater impact and durability than less thoughtfully generated judgments (Wegener, Petty, Blankenship, & Detweiler-Bedell, 2010), which means that high-performing students might be more likely to act on the basis of (or in response to) their predictions than low-performing students. Importantly, although the present evidence and some prior published evidence (i.e., Händel & Fritzsche, 2016; Miller & Geraci, 2011b) are consistent with the idea *that* low-performing people do not engage in as much effort to adjust their judgments from their desired grade anchor, these findings do not elucidate *why* low-performing people might not adjust their judgments.

Finally, it is well-established that people's JOLs are, at least in part, based on their interpretation of the available—and sometimes ambiguous—cues related to their learning (e.g., Koriat, 1997; Serra & Metcalfe, 2009). It is also well-established that people's desires have a powerful influence over perception, leading them to see ambiguous stimuli as those that they are looking for (Balcetis & Dunning, 2006) or see desired objects as closer (Balcetis & Dunning, 2010), along with a number of other, similar perceptual biases (for a review, see Dunning & Balcetis, 2013). For example, when judging their current level of mastery, students who want to perform well on an exam or in a course might overemphasize cues about their learning that are in line with the desires, and ignore cues that suggest that they have not yet mastered the materials. More relevant to the present situation, students' desires or goals might have biased their interpretation of the

metacognitive cues they used to make their performance predictions. In the present studies, students' desires to perform well on their exams might have led them to consider cues that supported the idea that they would perform well on the exam, leading to their optimistic performance predictions. The processes leading to this optimism, however, might not have led participants to be confident in that assessment. We do not yet know what factors influence participants' confidence in their metacognitive judgments (e.g., Dunlosky et al., 2005; Serra, Dunlosky, & Hertzog, 2008; Serra & England, 2012); identifying such factors certainly warrants empirical examination in the future.

### Closing remarks

To date, most discussion and study of the unskilled-and-unaware pattern has relied on the assumption that the pattern stems from at least two populations of people—the skilled and the unskilled—making performance predictions in two different ways. Unlike most current accounts of the unskilled-and-unaware pattern, the present account does not require the unskilled to be unaware of their abilities; instead, the present account utilizes the same mechanism—the influence of people's desires on their performance predictions—to explain the predictions of both the skilled and the unskilled. The present results have clear implications for explaining overconfidence in general, and the unskilled-and-unaware pattern specifically, because most students (including the lowest-performing ones) wanted to earn high grades on the exams, and subsequently demonstrated marked overconfidence in their grade predictions. Although the projection of students' desires is almost certainly not the only factor that drives the unskilled-and-unaware phenomenon, the present research suggests that these desires should be among the multitude of mechanisms considered when explaining this effect.

### References

- Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences*, *33*, 587–605. doi:10.1016/S0191-8869(01)00174-X
- Ariel, R., & Dunlosky, J. (2011). The sensitivity of judgment-of-learning resolution to past test performance, new learning, and forgetting. *Memory & Cognition*, *39*, 171–184. doi:10.3758/s13421-010-0002-y
- Balcetis, E., & Dunning, D. A. (2006). See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology*, *91*, 612–625. doi:10.1037/0022-3514.91.4.612
- Balcetis, E., & Dunning, D. (2010). Wishful seeing: More desired objects are seen as closer. *Psychological Science*, *21*, 147–152. doi:10.1177/0956797609356283
- Barden, J., & Petty, R. E. (2008). The mere perception of elaboration creates attitude certainty: Exploring the thoughtfulness heuristic. *Journal of Personality and Social Psychology*, *95*, 489–509. doi:10.1037/a0012559
- Billeter, D., Kalra, A., & Loewenstein, G. (2011). Underpredicting learning after initial experience with a product. *Journal of Consumer Research*, *37*, 723–736. doi:10.1086/655862
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444. doi:10.1146/annurev-psych-113011-143823
- Briñol, P., & DeMarree, K. G. (Eds.). (2012). *Social metacognition*. Hove, UK: Psychology Press.
- Buehler, R., Griffin, D., & Ross, M. (2002). Inside the planning fallacy: The causes and consequences of optimistic time predictions. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 251–270). Cambridge, UK: Cambridge University Press.
- Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology*, *90*, 60–77. doi:10.1037/0022-3514.90.1.60
- Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review*, *20*, 1350–1356. doi:10.3758/s13423-013-0442-z
- DeMarree, K. G., & Rios, K. (2014). Self-esteem and self-concept clarity: Exploring the role of desired self-esteem. *Journal of Experimental Social Psychology*, *50*, 202–209. doi:10.1016/j.jesp.2013.10.003
- DeMarree, K. G., Wheeler, S. C., Briñol, P., & Petty, R. E. (2014). Wanting other attitudes: Actual–desired attitude discrepancies predict feelings of ambivalence and ambivalence consequences. *Journal of Experimental Social Psychology*, *53*, 5–18. doi:10.1016/j.jesp.2014.02.001
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks, CA: Sage.
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*, 271–280. doi:10.1016/j.learninstruc.2011.08.003
- Dunlosky, J., Serra, M. J., Matvey, G., & Rawson, K. A. (2005). Second-order judgments about judgments of learning. *Journal of General Psychology*, *132*, 335–346. doi:10.3200/GENP.132.4.335-346
- Dunning, D., & Balcetis, E. (2013). Wishful seeing: How preferences shape visual perception. *Current Directions in Psychological Science*, *22*, 33–37. doi:10.1177/0963721412463693
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12*, 83–87. doi:10.1111/1467-8721.01235
- Ehrlinger, J., Johnson, K. L., Banner, M., Dunning, D. A., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, *105*, 98–121. doi:10.1016/j.obhdp.2007.05.002
- England, B. D., & Serra, M. J. (2012). The contribution of anchoring and past-test performance to the underconfidence-with-practice effect. *Psychonomic Bulletin & Review*, *19*, 715–722. doi:10.3758/s13423-012-0237-7
- Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review*, *101*, 519–527. doi:10.1037/0033-295X.101.3.519
- Finn, B. (2008). Framing effects on metacognitive monitoring and control. *Memory & Cognition*, *36*, 813–821. doi:10.3758/MC.36.4.813
- Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental*

- Psychology: Learning, Memory, and Cognition*, 33, 238–244. doi:10.1037/0278-7393.33.1.238
- Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, 15, 84–93. doi:10.3758/BF03197714
- Griffin, T. D., Jee, B. D., & Wiley, J. (2009). The effects of domain knowledge on metacomprehension accuracy. *Memory & Cognition*, 37, 1001–1013. doi:10.3758/MC.37.7.1001
- Grimes, P. W. (2002). The overconfident principles of economics student: An examination of a metacognitive skill. *Journal of Economic Education*, 33, 15–30. doi:10.1080/00220480209596121
- Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: the effects of incentives, reflection, and explanatory style. *Metacognition and Learning*, 3, 101–121. doi:10.1007/s11409-008-9021-5
- Händel, M., & Fritzsche, E. S. (2016). Unskilled but subjectively aware: Metacognitive monitoring ability and respective awareness in low-performing students. *Memory & Cognition*, 44, 229–241. doi:10.3758/s13421-015-0552-0
- Hartwig, M. K., & Dunlosky, J. (2014). The contribution of judgment scale to the unskilled-and-unaware phenomenon: How evaluating others can exaggerate over- (and under-) confidence. *Memory & Cognition*, 42, 164–173. doi:10.3758/s13421-013-0351-4
- Klayman, J., Soll, J. B., González-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes*, 79, 216–247. doi:10.1006/obhd.1999.2847
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349–370. doi:10.1037/0096-3445.126.4.349
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 107–118. doi:10.1037/0278-7393.6.2.107
- Kornell, N., & Bjork, R. A. (2008). Optimising self-regulated study: The benefits—and costs—of dropping flashcards. *Memory*, 16, 125–136. doi:10.1080/09658210701763899
- Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 609–622. doi:10.1037/0278-7393.32.3.609
- Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology*, 82, 180–188. doi:10.1037/0022-3514.82.2.180
- 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, 1121–1134. doi:10.1037/0022-3514.77.6.1121
- Kruger, J., & Dunning, D. (2002). Unskilled and unaware—but why? A reply to Krueger and Mueller. *Journal of Personality and Social Psychology*, 82, 189–192. doi:10.1037/0022-3514.82.2.189
- Lipko, A. R., Dunlosky, J., & Merriman, W. E. (2009). Persistent overconfidence despite practice: The role of task experience in preschoolers' recall predictions. *Journal of Experimental Child Psychology*, 103, 152–166. doi:10.1016/j.jecp.2008.10.002
- Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in theory and practice* (pp. 117–144). Mahwah, NJ: Erlbaum.
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, 15, 174–179. doi:10.3758/PBR.15.1.174
- Miller, T. M., & Geraci, L. (2011a). Training metacognition in the classroom: The influence of incentives and feedback on exam predictions. *Metacognition and Learning*, 6, 303–314. doi:10.1007/s11409-011-9083-7
- Miller, T. M., & Geraci, L. (2011b). Unskilled but aware: Reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 502–506. doi:10.1037/a0021802
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125–173). New York, NY: Academic Press.
- Scheck, P., Meeter, M., & Nelson, T. O. (2004). Anchoring effects in the absolute accuracy of immediate versus delayed judgments of learning. *Journal of Memory and Language*, 51, 71–79. doi:10.1016/j.jml.2004.03.004
- Schneider, W. (1998). Performance prediction in young children: Effects of skill, metacognition and wishful thinking. *Developmental Science*, 1, 291–297. doi:10.1111/1467-7687.00044
- Serra, M. J., Dunlosky, J., & Hertzog, C. (2008). Do older adults show less confidence in their monitoring of learning? *Experimental Aging Research*, 34, 379–391. doi:10.1080/03610730802271898
- Serra, M. J., & England, B. D. (2012). Magnitude and accuracy differences between judgments of remembering and forgetting. *Quarterly Journal of Experimental Psychology*, 65, 2231–2257. doi:10.1080/17470218.2012.685081
- Serra, M. J., & Metcalfe, J. (2009). Effective implementation of metacognition. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition and education* (pp. 278–298). New York, NY: Routledge.
- Shanks, L. L., & Serra, M. J. (2014). Domain familiarity as a cue for judgments of learning. *Psychonomic Bulletin & Review*, 21, 445–453. doi:10.3758/s13423-013-0513-1
- Soderstrom, N. C., & McCabe, D. P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: Evidence for agenda-based monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1236–1242. doi:10.1037/a0023548
- Stinebrickner, R., & Stinebrickner, T. R. (2014). A major in science? initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies*, 81, 426–472. doi:10.1093/restud/rdt025
- Stipek, D. J., Roberts, T. A., & Sanborn, M. E. (1984). Preschool-age children's performance expectations for themselves and another child as a function of the incentive value of success and the salience of past performance. *Child Development*, 55, 1983–1989. doi:10.2307/1129773
- Tauber, S. K., & Rhodes, M. G. (2012). Multiple bases for young and older adults' judgments-of-learning (JOLs) in multitrial learning. *Psychology and Aging*, 27, 474–483. doi:10.1037/a0025246
- Wegener, D. T., Petty, R. E., Blankenship, K. L., & Detweiler-Bedell, B. (2010). Elaboration and numerical anchoring: Implications of attitude theories for consumer judgment and decision making. *Journal of Consumer Psychology*, 20, 5–16. doi:10.1016/j.jcps.2009.12.003
- Willard, G., & Gramzow, R. H. (2009). Beyond oversights, lies, and pies in the sky: Exaggeration as goal projection. *Personality and Social Psychology Bulletin*, 35, 477–492. doi:10.1177/0146167208329631
- Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, 27, 551–572. doi:10.1016/S0361-476X(02)00006-1