Educational Psychology Review
https://doi.org/10.1007/s10648-021-09598-6

INTERVENTION STUDY

# Test Anxiety and Metacognitive Performance in the Classroom

**Katie M. Silaj**[1] · **Shawn T. Schwartz**[1,2] · **Alexander L. M. Siegel**[1,3] · **Alan D. Castel**[1]

Accepted: 26 January 2021/Published online: 18 March 2021


**Abstract**
Test anxiety is a context-specific academic anxiety which can result in poorer academic and metacognitive performance. We assessed how the quantity and relative weight of assessments contribute to the effects of test anxiety on performance and metacognitive accuracy in a smaller seminar-style class on human memory (study 1) and a larger lecture-style class on cognitive psychology (study 2). Students took six low-stakes quizzes each worth 10% of their final grade in study 1 and two high-stakes exams each worth 40% of their final grade in study 2. All students provided their state anxiety and predicted their scores before and after each assessment. Students in both classes also provided their trait (overall) anxiety after the final assessment. In both studies, students' higher post-state anxiety appeared to be associated with worse assessment performance; however, pre- and post-state anxiety decreased across the quarter in study 1 but remained constant in study 2. Additionally, we found that metacognitive accuracy moderated the effect of post-state anxiety on performance in study 1. Students with higher trait anxiety in study 1 were underconfident in their scoring predictions, while in study 2 students with higher trait anxiety performed worse on their assessments. Thus, students' metacognitive accuracy appears to be influenced by trait anxiety when taking low-stakes quizzes, while performance is related to trait anxiety when taking high-stakes exams.

**Keywords** Test anxiety · Metacognition · Trait anxiety · Low-stakes testing · Classroom study

✉ Katie M. Silaj
kmsilaj@ucla.edu

1 Department of Psychology, University of California, Los Angeles, 502 Portola Plaza, Los Angeles, CA 90095, USA

2 Present address: Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA, USA

3 Present address: Leonard Davis School of Gerontology, University of Southern California, Los Angeles, CA, USA

Test anxiety is a context-specific academic anxiety in which an individual reacts to an evaluative event like a quiz, midterm test, or final exam in an anxious, fearful, or nervous manner (Cassady, 2010). It has been estimated to impact 25–40% of students (Cassady, 2010) and can result in a combination of negative outcomes inducing poor test performance (Cassady & Johnson, 2002). Test anxiety can be viewed as both a state anxiety that is triggered by the testing event, but also as a trait anxiety in which the effects of test anxiety are more severe for people with high levels of baseline anxiety (*hereafter* "trait anxiety"; Goetz et al., 2013; Harris et al., 2019; Spielberger, 1972). According to Zohar (1998), the impact of test anxiety depends on both the individual's trait anxiety and the situational factors surrounding the test, such as the individual's level of confidence in their content knowledge or the level of stakes associated with the test.

Interventions for test anxiety can increase students' performance while having no effect on trait anxiety (Harris et al., 2019). Thus, it is important to consider individual differences in trait anxiety when investigating the impact of test anxiety on performance.

Furthermore, external factors may also play a role in the effects of test anxiety, such as the stakes (Wood et al., 2016) and format (Hembree, 1988) of the assessment. For example, higher test anxiety is associated with poor performance on high-stakes standardized tests (Cassady & Johnson, 2002; Putwain, 2008), and the multiple-choice testing format is associated with higher test anxiety when compared to a matching format (Hembree, 1988) or an open-ended format (Birenbaum & Feldman, 1998).

Lastly, *metacognition*, or the ability to assess and control our cognitive processes (Everson et al., 1994), is related to test anxiety such that students with high test anxiety often show a lower level of metacognitive skillfulness compared to students with low test anxiety. Some work has suggested that metacognition may be a mediator between test anxiety and performance (Veenman et al., 2000) and between test anxiety and study strategies (Spada et al., 2006). Hence, the focus of the current study was to examine how self-reported state anxiety affects student performance and metacognitive accuracy, and how those measures vary with levels of self-reported trait anxiety along with the number and weight of assessments students take in a given course.

## Test Anxiety and Academic Performance

Test anxiety is associated with poorer academic performance in elementary, high school, and college students (Cassady & Johnson, 2002; Putwain & Best, 2011; Williams, 1991) and is considered to be one of the most disruptive factors in exam performance (Cizek & Burg, 2006; von der Embse & Hasson, 2012). High test anxiety is related to lower test scores in general and, specifically, to poor performance on college entrance exams and standardized tests, which provides evidence that the stakes of assessments may induce higher levels of test anxiety (von der Embse et al., 2018). In one study, von der Embse and Hasson (2012) examined the effects of test anxiety on student performance on a high-stakes state graduation exam in an urban and a suburban high school. They found that test anxiety accounted for 4–15% of the variance in test scores when controlling for school. This result illustrates the wide-ranging effects of test anxiety and raises the concern that the effects of test anxiety and poor performance on high-stakes assessments may be compounded for students in school districts that fail to meet standards for funding. Additionally, poor performance on high-stakes exams could lead to other negative consequences for students such as low course grades (Segool et al., 2013), less attentional control during exams (Fernández-Castillo & Caurcel, 2015), and increased anxiety and depression (Leadbeater et al., 2012; von der Embse et al., 2018). Thus, it is important to better understand which factors associated with real classroom environments are related to the relationship between test anxiety and performance.

## Interventions for Test Anxiety

Despite decades of research on test anxiety, there remains a lack of consensus amongst experts in the field regarding the most effective interventions to combat its detrimental effects on student performance. One intervention that an instructor might implement to encourage the development of better study habits and a more positive outlook on testing events is to frame assessments as "retrieval practice." There is a considerable amount of research that supports the idea that testing oneself is a much more powerful learning event than simply restudying material (Roediger & Karpicke, 2006), with retrieval practice being described as a desirable difficulty (Bjork, 1994). The positive implications of testing oneself are often not apparent until after the learner has taken the assessment and may seem like a challenge to the learner initially. This knowledge has led researchers in applied educational settings to implement testing as an intervention to promote student retention of learned content (Karpicke, 2017).

The results of these classroom experiments have led to recommendations that teachers should use testing as a tool to further student learning (Dunlosky et al., 2013; Dunn et al., 2013). Perhaps regarding test anxiety, more frequent testing events provided by the instructor can help the learner to engage in retrieval practice, which may negate the feelings of anxiousness in response to tests over time. A recent study investigated the effects of retrieval practice on test anxiety in middle and high school students (Argarwal et al., 2014) and reported that 72% of students credited retrieval practice with making them feel less nervous for tests and exams. As such, this may be a promising avenue for reducing test anxiety in a classroom setting, although it has yet to have been explored within a real-world college setting.

An intervention that may be more easily implemented in an undergraduate education context is varying the frequency of testing events within a given course. In a recent study, participants studied five word lists, and then after a distractor task, they either took an interim test or restudied the words (Yang et al., 2020). Performance on a cumulative test at the end of the study was significantly higher for participants who took interim tests after each list compared with the group that restudied the words. Additionally, participants took a Test Anxiety Inventory (TAI; Chinese version; Yue, 1996) which revealed that the low-stakes assessments can improve performance regardless of levels of test anxiety. Although this was demonstrated in a controlled, experimental context, we are currently interested in examining the effects of format and relative weight of assessments on performance and metacognitive accuracy in a classroom setting.

## Test Anxiety and Metacognitive Accuracy

Test anxiety and metacognition are intimately linked (Spada et al., 2006). Thus, the process of monitoring oneself throughout the testing cycle may be an important factor in determining performance outcomes. Heightened anxiety brought upon by metacognitive awareness of a lack of preparation or ability causes the learner to become more metacognitively aware of their failure to succeed on the test, thus inducing more anxiety (Covington, 1985; Naveh-Benjamin, 1991). This can lead to additional cognitive interference because these resources that could have been devoted to retrieving test-relevant knowledge are essentially divided due to heightened levels of anxiety (Cassady & Johnson, 2002; Kurosawa & Harackiewicz, 1995; Veenman et al., 2000). Furthermore, individual differences exist such that students with high test anxiety may display lower levels of metacognitive abilities when compared to students with low levels of test anxiety (Veenman et al., 2000).

Veenman et al. (2000) assessed this relationship between metacognition and performance using the metacognitive word knowledge task. In this task, participants were asked if they knew the definitions of 31 health science words, and then were asked to provide definitions. The authors assessed metacognition by computing hits (e.g., the number of words participants said they knew and correctly identified their meaning on the vocabulary task) and false alarms (e.g., the number of words participants said they knew but did not correctly identify their meaning on the vocabulary task) rates. Participants with low test anxiety exhibited a superior metacognitive skillfulness during math performance relative to participants with high test anxiety (Veenman et al., 2000). In another study, test anxiety exerted a negative influence on students' performance on a metacognitive word knowledge task, independent of overall reading ability (Everson et al., 1994).

Metacognitive accuracy can interact with math anxiety such that higher math anxiety predicts lower test performance except in students that are high in metacognitive skillfulness (Legg & Locker, 2009). Additionally, when students are more metacognitively aware, they report higher confidence in their ability to answer test problems correctly. Some work has shown that with feedback, students can improve their metacognitive accuracy from one exam to another (Callender et al., 2016). Thus, more retrieval practice opportunities within an academic term may lead to increased metacognition, which may especially benefit students that are high in test anxiety.

## Overview of the Present Study

In the current study, we examined the effects of self-reported state anxiety on performance and metacognitive accuracy in a true classroom setting, and how these measures vary with students' levels of self-reported trait anxiety and (i) the number and (ii) relative weight of assessments on a student's overall grade within a given course. In a smaller, seminar-style class in human memory, students took six quizzes that were each worth 10% of their final grade (study 1), while in a larger, lecture-style class in cognitive psychology, students took two exams, each worth 40% of their grade (study 2). Thus, study 1 provided a context in which we could examine the impact of state and trait anxiety on student performance and metacognition in a smaller, low-stakes setting where students had more opportunities for retrieval practice, while study 2 provided a context in which we could examine the impact of state and trait anxiety on student performance and metacognition in a larger, high-stakes environment with less retrieval practice opportunities.

Students in both classes completed a questionnaire both before and after each assessment within the span of the academic quarter (11 weeks: 10 instructional weeks + 1 final exam week). The questionnaires prompted students to predict and postdict their scores on the assessment and rate their level of state anxiety pre- and post-assessment. Students also gave a rating of trait anxiety either after the final assessment of the academic quarter (study 1) or after each assessment (study 2). We used single-report measures as opposed to a formal questionnaire to avoid introducing additional anxiety into the testing environment and as a practical means to address time constraints imposed by the classroom setting. We were interested in measuring self-reported state anxiety and its relationship with performance and metacognitive accuracy while controlling for individual differences in self-reported trait anxiety.

We expected that student scores on assessments and their metacognitive accuracy in predicting their scores would significantly increase across the quarter for both classes due to experience with the content and expectations of the course. We also expected state anxiety to significantly decrease from pre- to post-assessment for students in both classes due to a sense of relief that the test is over, in addition to across the quarter for students who engaged in more frequent, relatively lower-stakes assessments as they will gain retrieval practice (Argarwal et al., 2014). We expected students who report high ratings of trait anxiety to also report significantly higher ratings of state anxiety and have significantly lower scores on the assessments compared to students who report low ratings of trait anxiety (Zohar, 1998). Lastly, we expected metacognition to significantly interact with test anxiety, such that students with high test anxiety would be less accurate in their assessment score predictions and postdictions compared to students with low test anxiety (Veenman et al., 2000).

## Study 1

### Method

#### Participants

Participants were 49 University of California, Los Angeles (UCLA), undergraduate students enrolled in Psychology 124C: Human Memory during Winter Quarter 2018. One outlier was removed due to providing ratings outside of the requested range, so there were 48 students included in the final sample.

#### Materials and Procedure

There were six, short-answer format, low-stakes quizzes given during weeks 3–9 (no quiz week 6). The quizzes covered materials from the lecture and assigned readings from the previous week. Possible quiz scores ranged from 0 to 10 and the lowest quiz score was dropped at the end of the quarter. Quizzes accounted for 50% of the final grade (each quiz was worth 10%). There was also a written assignment, research proposal, and in-class presentation that in combination contributed the remaining 50% of the final grade (not included in the data analyzed here). Quiz topics included (*in this order*) short-term and working memory, encoding and studying, forgetting and retrieval/testing, autobiographical and eyewitness memory, self-regulated learning, metamemory, memory and aging, and expertise and applications. An example of the type of question a student may have answered on a quiz would be of the following form and level of conceptual understanding: *"Define and distinguish between proactive and retroactive interference and give unique examples of each. How can you show a release of proactive interference, and what does this suggest about the nature of forgetting?"*

Before each quiz, students were asked *"How nervous/anxious do you feel at the current moment?"* and would respond on a Likert scale from 1 (*not at all anxious*) to 10 (*very much anxious*). We used this measure to assess student levels of state anxiety in an effort to minimize any effects of anxiety a full questionnaire might induce. Single-scale measures have been found to correlate well with more extensive scales (Davey et al., 2007; Núñez-Peña et al., 2013) and are more practical and less invasive for our purposes when considering inherent student anxiety and time constraints in a more naturalistic, classroom setting.

Students were then asked to predict their scores out of 10 points. They had 45 min to answer the three short answer questions. After the quiz, students reported their state anxiety on the same scale a second time. Lastly, they were asked to postdict their scores out of 10 points. The students were told that their responses to the pre- and post-quiz surveys were voluntary and would not influence their grades. This procedure was followed for all six quizzes. At the end of quiz 6 (week 9), students were asked *"In general, how anxious of a person do you think you are?"* and responded on a Likert scale ranging from 1 (*not at all anxious in general*) to 10 (*very much anxious in general*). This measure was requested following the final post-quiz state anxiety rating of the academic quarter and was used to assess individual levels of trait anxiety.

Additionally, an assigned course "reader" graded all assignments for the course and was blind to the hypotheses and predictions of the current study as well as unaware of participants' self-reported ratings when grading the quizzes. All materials and procedures were performed ethically and approved by the UCLA Institutional Review Board.

## Results and Discussion

For the following analyses, we first conducted a repeated-measures analysis of variance (ANOVA) on each dependent variable while treating quiz number throughout the quarter as a within-subjects independent measure (for ANOVAs where Mauchly's test of sphericity was violated, Greenhouse-Geisser sphericity corrections were used). Then, we conducted both multiple regression analyses and interrupted regressions (Simonsohn, 2018) to test if quiz number significantly predicted participants' overall quiz performance, estimation accuracy, and anxiety throughout the quarter. We did not assume monotonicity of each relationship, and therefore tested the plausibility of both the presence of linear and non-linear relationships between quiz number and each dependent variable (Bowman et al., 1998). Interrupted regressions were subsequently used to reveal the potential existence of U-shaped (or inverted U-shaped) relationships and were conducted using the "Robin Hood" algorithm to identify a break point along measures of the dependent variable (two-lines test, Simonsohn, 2018). Upon examining dependent variables throughout the academic quarter, a U-shaped relationship might have occurred if, for example, a dependent measure was higher or lower at both the beginning and end of the academic quarter relative to the middle. Such a pattern may result from external factors potentially elevating levels of anxiety or performance due to the nature of uncertainty when becoming acquainted with a new course at the beginning of a term or knowing that stakes are high for students with borderline grades at the end of the term. Additionally, the interrupted regression model using the "Robin Hood" algorithm generally yields higher statistical power for detecting U-shaped relationships than other statistical alternatives, and also is less likely to result in higher rates of false-positive and false-negative results that can occur when assuming and forcing a standard quadratic functional form on the data (Simonsohn, 2018). In accordance with Simonsohn (2018), we deemed a U-shaped relationship to exist, and therefore rejected the null hypothesis of an absence of a U-shaped relationship, if both slopes of the two lines were significant and of opposite sign. For each dependent variable, if both the results from the multiple regression model and the interrupted regression model were significant, we opted for the most parsimonious explanation by choosing the multiple regression (linear) model to explain the apparent relationship between the data. Additionally, we reported adjusted $R^2$ values as percent of variance explained for all linear regressions.

## Quiz Performance Throughout the Quarter

To examine overall quiz performance with respect to quiz number, we conducted a 1 (Exam: *quiz*) × 6 (Quiz number: *1, 2, ..., 6*) repeated-measures ANOVA on overall quiz scores, which revealed a significant main effect of quiz number, $F(2.78, 108.50) = 10.03$, $p < .001$, $\eta^2 = .20$ (Table 1). Post hoc paired-samples *t*-tests with a Bonferroni correction indicated that scores were higher on quizzes 3 ($M = 8.58$, $SD = 1.31$), 4 ($M = 9.35$, $SD = .62$), 5 ($M = 8.85$, $SD = 1.03$), and 6 ($M = 8.98$, $SD = 1.07$) relative to quiz 1 ($M = 7.88$, $SD = 1.91$), higher on quiz 4 ($M = 9.35$, $SD = .62$) relative to quiz 2 ($M = 8.45$, $SD = .95$), and higher on quiz 4 ($M = 9.35$, $SD = .62$) relative to quiz 3 ($M = 8.58$, $SD = 1.31$), adjusted *p*s < .03. No other comparisons were significant, adjusted *p*s > .17.

Next, we conducted regressions using both pre-quiz and post-quiz state anxiety as predictors. The multiple regression model took the following form: quiz score = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (pre-quiz state anxiety) + $\beta_3$ (post-quiz state anxiety) + $\beta_4$ (trait anxiety). Four predictors significantly explained 14.8% of the variance, $R^2 = .16$, $F(4, 269) = 12.84$, $p < .0001$, and we found that quiz number ($\beta_1 = .20$, $p < .0001$) and post-quiz state anxiety ($\beta_3 = -.19$, $p < .0001$) significantly predicted quiz score, but pre-quiz state anxiety ($\beta_2 = .05$, $p = .24$) and trait anxiety ($\beta_4 = .04$, $p = .27$) did not (Table 2). The interrupted regression model took the same form as the multiple regression model and revealed no significant U-shaped relationship $\beta_{1–4} = .38$, $p < .0001$; $\beta_{5–6} = -.12$, $p = .24$). Overall, these findings indicate a significant positive linear relationship between quiz number and average quiz score, such that when controlling for both pre-quiz and post-quiz state anxiety along with trait anxiety, the average quiz score for each student increased with each subsequent quiz throughout the quarter (see Fig. 1).

**Table 1** Repeated-measures analysis of variance coefficients in study 1

| Outcome | *F* | *df* | *p* | $\eta^2$ |
|---|---|---|---|---|
| Quiz performance | 10.03 | (2.78, 108.50) | < .001*** | .20 |
| Predictions | 4.71 | (3.70, 136.95) | .002** | .11 |
| Postdictions | 7.62 | (3.17, 104.74) | < .001*** | .19 |
| Prediction calibration | 2.01 | (5, 195) | .08 | .05 |
| Postdiction calibration | 1.93 | (3.40, 132.55) | .12 | .05 |
| Overall calibration | 2.66 | (5, 195) | .02* | .06 |
| Calibration difference score | 4.12 | (3.62, 170.26) | .004** | .08 |
| Pre-quiz state anxiety | 4.66 | (2.21, 10.26) | < .001*** | .11 |
| Post-quiz state anxiety | 3.16 | (5, 170) | .009** | .09 |
| State anxiety difference score | 3.81 | (5, 235) | .002** | .08 |

*$p < .05$, **$p < .01$, ***$p < .001$

**Table 2** Multiple regression analyses coefficients using pre-quiz and post-quiz state anxiety as a predictor in study 1

| Outcome | % variance explained Explained | $R^2$ | $F$ | $df$ | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|---|---|---|---|---|---|
| Quiz score | 14.8% | .16**** | 12.84 | (4, 269) | 8.32**** | .20**** | .05 | −.19**** | .04 |
| Prediction | 5.7% | .07*** | 5.13 | (4, 267) | 8.33**** | .05 | −.11* | −.02 | −.09* |
| Postdiction | 18.5% | .20**** | 16.39 | (4, 267) | 9.54**** | −.02 | .09 | −.29**** | −.11** |
| Prediction calibration | 8.0% | .09**** | 6.90 | (4, 269) | −.22 | .17* | .16* | −.16* | .15** |
| Postdiction calibration | 11.2% | .12**** | 9.57 | (4, 269) | −1.20** | .19*** | −.03 | .13* | .15*** |
| Overall calibration | 9.1% | .10**** | 7.81 | (4, 269) | −.71 | .18** | .06 | −.02 | .15*** |
| Calibration difference score | 7.1% | .08**** | 6.25 | (4, 269) | 1.16*** | −.13** | .13** | −.12** | .04 |

*$p$ < .05, **$p$ < .01, ***$p$ < .001, ****$p$ < .0001

Model: Outcome = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (pre-quiz state anxiety) + $\beta_3$ (post-quiz state anxiety) + $\beta_4$ (trait anxiety)

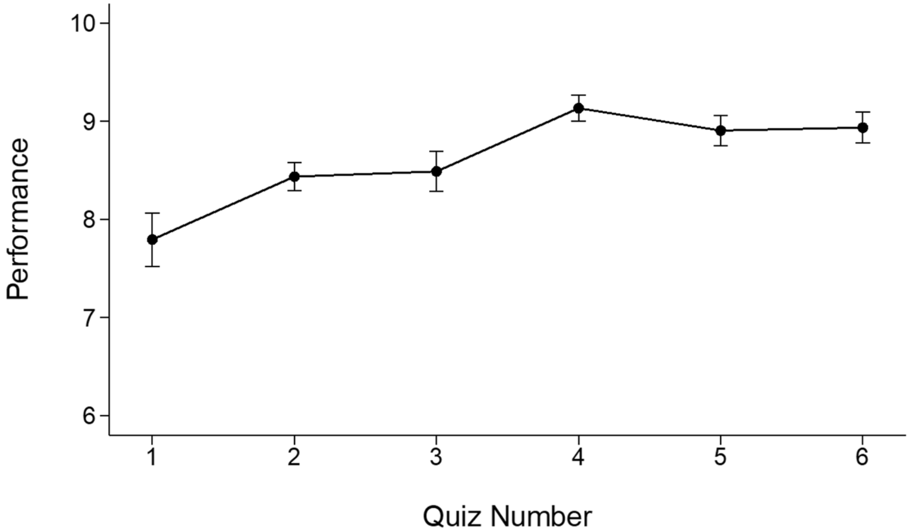


**Fig. 1** Mean quiz score performance in study 1 across the academic quarter. All error bars represent ± 1 standard error

## Estimation Accuracy Throughout the Quarter

To examine quiz score metacognitive prediction (prior to taking each quiz) with respect to quiz number, we conducted a 1 (Exam: *quiz*) × 6 (Quiz number: *1, 2, ..., 6*) repeated-measures ANOVA on quiz score prediction, which revealed a significant main effect of quiz number, $F(3.70, 136.95) = 4.71$, $p = .002$, $\eta^2 = .11$ (Table 1). Post hoc paired-samples *t*-tests with a Bonferroni correction indicated that metacognitive score prediction was higher on quiz 4 ($M = 7.87$, $SD = 1.25$) than on quiz 1 ($M = 6.96$, $SD = 1.94$), and lower on quiz 5 ($M = 7.07$, $SD = 1.53$) than on quiz 4 ($M = 7.87$, $SD = 1.25$), adjusted *p*s < .006. No other comparisons were significant, adjusted *p*s > .055.

Next, we conducted separate regressions examining quiz score prediction and postdiction, each using both pre-quiz state anxiety and post-quiz state anxiety as predictors. For quiz score prediction, the multiple regression model took the following form: quiz score prediction = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (pre-quiz state anxiety) + $\beta_3$ (post-quiz state anxiety) + $\beta_4$ (trait anxiety). The four predictors significantly explained 5.7% of the variance, $R^2 = .07$, $F(4, 267) = 5.13$, $p = .0005$, and we found that pre-quiz state anxiety significantly predicted quiz score prediction ($\beta_2 = -.11$, $p = .04$) as did trait anxiety ($\beta_4 = -.09$, $p = .04$), but quiz number ($\beta_1 = .05$, $p = .36$) and post-quiz state anxiety ($\beta_3 = -.02$, $p = .76$) did not. The interrupted regression model took the same form as the multiple regression model and revealed no significant U-shaped relationship ($\beta_{1–3} = .31$, $p = .05$; $\beta_{4–6} = -.10$, $p = .29$) (Table 2). Thus, while there appeared to be a main effect of quiz number, when controlling for participants' state anxiety and trait anxiety, quiz number was no longer a significant predictor of quiz score prediction (see Fig. 2a).

Secondly, to examine quiz score metacognitive postdiction (directly following taking each quiz) with respect to quiz number, we conducted a 1 (Exam: *quiz*) × 6 (Quiz number: *1, 2,..., 6*) repeated-measures ANOVA on quiz score postdiction, which revealed a significant main effect of quiz number, $F(3.17, 104.74) = 7.62$, $p < .001$, $\eta^2 = .19$ (Table 1). Post hoc paired-

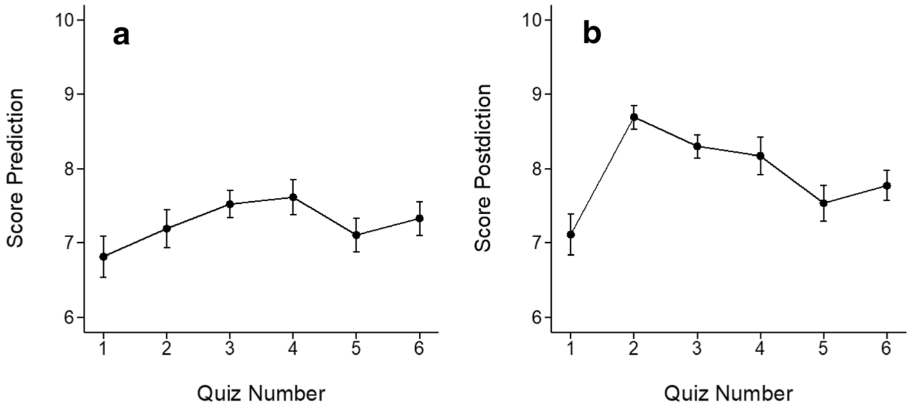


**Fig. 2** Mean quiz score predictions (**a**) and postdictions (**b**) of performance in study 1 across the academic quarter. All error bars represent ± 1 standard error

samples *t*-tests with a Bonferroni correction indicated that metacognitive score postdiction was higher on quizzes 2 ($M = 8.71$, $SD = 1.02$), 3 ($M = 8.25$, $SD = 1.10$), and 4 ($M = 8.32$, $SD = 1.79$) relative to quiz 1 ($M = 7.37$, $SD = 1.63$), and lower on quiz 5 ($M = 7.65$, $SD = 1.70$) compared to quiz 2 ($M = 8.71$, $SD = 1.02$), adjusted *p*s < .008. No other comparisons were significant, adjusted *p*s > .10.

For quiz score postdiction, the multiple regression model took the following form: quiz score postdiction = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (pre-quiz state anxiety) + $\beta_3$ (post-quiz state anxiety) + $\beta_4$ (trait anxiety). The four predictors significantly explained 18.5% of the variance, $R^2 = .20$, $F(4, 267) = 16.39$, $p < .0001$, and we found that post-quiz state anxiety significantly predicted quiz score postdiction ($\beta_3 = -.29$, $p < .0001$) as did trait anxiety ($\beta_4 = -.11$, $p = .005$), but quiz number $\beta_1 = -.02$, $p = .63$) and pre-quiz state anxiety ($\beta_2 = .09$, $p = .08$) did not. The interrupted regression model took the same form as the multiple regression model and revealed a significant inverted U-shaped relationship ($\beta_{1-2} = 1.34$, $p < .0001$; $\beta_{3-6} = -.23$, $p < .0001$) (Table 2). Thus, while there appeared to be a main effect of quiz number, controlling for participants' state anxiety and trait anxiety revealed that quiz number was no longer a significant linear predictor of quiz score postdiction; however, quiz score postdiction was seemingly lower at both the beginning and end of the academic quarter relative to the middle as revealed by the significant inverted U-shaped relationship (see Fig. 2b).

We then examined another form of estimation accuracy using students' pre- and postdiction calibration (*actual score – predicted/postdicted score*; such that a positive value would indicate underconfidence and a negative value would indicate overconfidence). To examine quiz score prediction estimation accuracy (for metacognitive predictions provided prior to taking each quiz) with respect to quiz number, we conducted a 1 (Exam: *quiz*) × 6 (Quiz number: *1, 2, ..., 6*) repeated-measures ANOVA on quiz score prediction calibrations, which did not reveal a significant main effect of quiz number, $F(5, 195) = 2.01$, $p = .08$, $\eta^2 = .05$ (Table 1).

For quiz score prediction calibration, the multiple regression model took the following form: quiz score prediction calibration = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (pre-quiz state anxiety) + $\beta_3$ (post-quiz state anxiety) + $\beta_4$ (trait anxiety). The four predictors significantly explained 8.0% of the variance, $R^2 = .09$, $F(4, 269) = 6.90$, $p < .0001$, and we found that quiz number significantly predicted quiz score prediction calibration ($\beta_1 = .17$, $p = .01$), as did pre-quiz state anxiety ($\beta_2 = .16$, $p = .02$), post-quiz state anxiety ($\beta_3 = -.16$, $p = .01$), and trait anxiety ($\beta_4 =$

.15, $p = .003$). The interrupted regression model took the same form as the multiple regression model and revealed no significant U-shaped relationship ($\beta_{1–5} = .20$, $p = .03$; $\beta_6 = -.01$, $p = .97$) (Table 2). Overall, these findings indicate a significant positive linear relationship for quiz score prediction calibration through time, as when controlling for both state and trait anxiety, the average quiz score prediction calibration for each student increased with each subsequent quiz throughout the quarter (see Fig. 3a).

Secondly, to examine quiz score postdiction estimation accuracy (for metacognitive postdictions provided directly following taking each quiz) with respect to quiz number, we conducted a 1 (Exam: *quiz*) × 6 (Quiz number: *1, 2, …, 6*) repeated-measures ANOVA on quiz score postdiction calibration, which did not reveal a significant main effect of quiz number, $F(3.40, 132.55) = 1.93$, $p = .12$, $\eta^2 = .05$ (Table 1).

For quiz score postdiction calibration, the multiple regression model took the following form: quiz score postdiction calibration = $\beta_0$ + $\beta_1$ (quiz number) + $\beta_2$ (pre-quiz state anxiety) + $\beta_3$ (post-quiz state anxiety) + $\beta_4$ (trait anxiety). The four predictors significantly explained

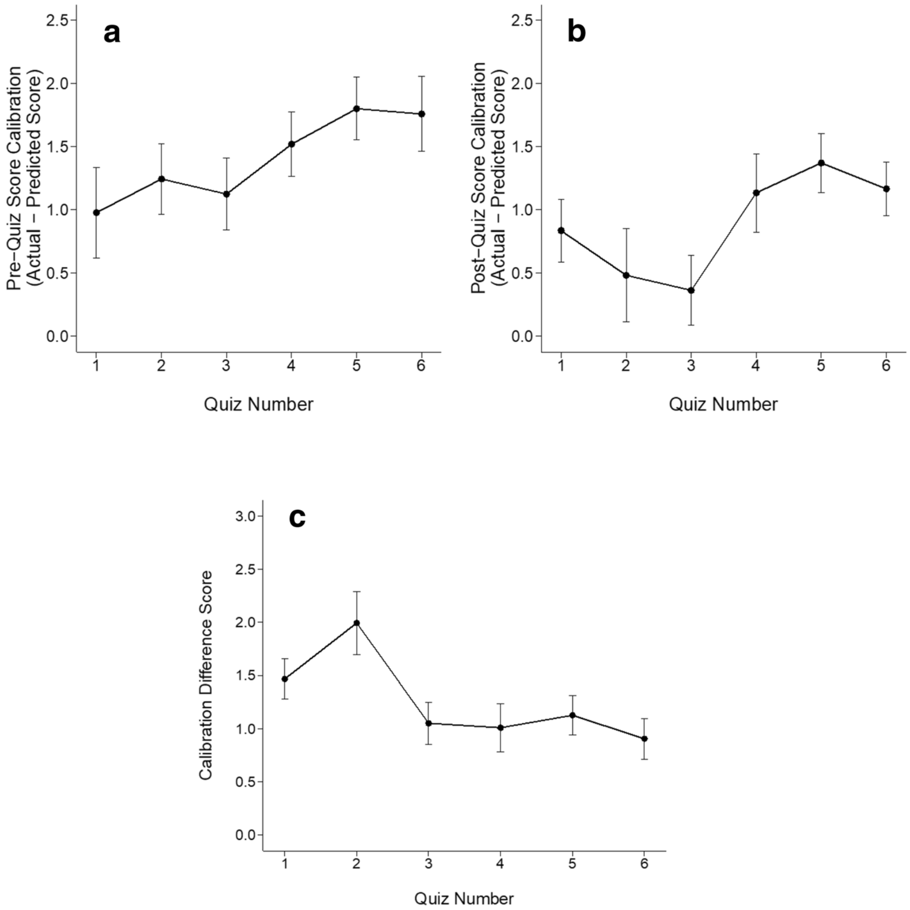


**Fig. 3** Mean score calibration (*actual − pre-/postdicted score*) before (pre-quiz; panel **a**) and after (post-quiz; panel **b**) each quiz in study 1 across the academic quarter. Mean calibration difference scores (*absolute value of difference between pre-quiz and post-quiz score calibration*; panel **c**) in study 1 across the academic quarter. All error bars represent ± 1 standard error

11.2% of the variance, $R^2 = .12$, $F(4, 269) = 9.57$, $p < .0001$, and we found that quiz number significantly predicted quiz score postdiction calibration ($\beta_1 = .19$, $p = .0005$), as did post-quiz state anxiety ($\beta_3 = .13$, $p = .01$) and trait anxiety ($\beta_4 = .15$, $p = .0005$), but pre-quiz state anxiety ($\beta_2 = -.03$, $p = .53$) did not. The interrupted regression model took the same form as the multiple regression model and revealed no significant U-shaped relationship ($\beta_{1-3} = -.25$, $p = .15$; $\beta_{4-6} = .31$, $p = .001$) (Table 2). Similarly, these findings indicate a significant positive linear relationship for quiz score postdiction calibration through time, as when controlling for both state and trait anxiety, the average quiz score postdiction calibration for each student increased with each subsequent quiz throughout the quarter (see Fig. 3b).

Additionally, to test whether or not students' overall calibration (the average of the prediction and postdiction score calibration) was generally underconfident or overconfident, we conducted a one-sample *t*-test with the alternative hypothesis that students' average overall calibration would be greater than 0 (i.e., underconfident), and found this to be the case, $t(47) = 6.71$, $p < .001$, Cohen's $d = .97$. We then conducted a 1 (Exam: *quiz*) × 6 (Quiz number: *1, 2, ..., 6*) repeated-measures ANOVA on students' overall calibration, which revealed a significant main effect of quiz number, $F(5, 195) = 2.66$, $p = .02$, $\eta^2 = .06$ (Table 1). Post hoc paired-samples *t*-tests with a Bonferroni correction indicated no significant differences for overall quiz score calibration between any other quiz, adjusted *p*s > .25. Next, we used a multiple regression model that took the following form: quiz score overall calibration = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (pre-quiz state anxiety) + $\beta_3$ (post-quiz state anxiety) + $\beta_4$ (trait anxiety). The four predictors significantly explained 9.1% of the variance, $R^2 = .10$, $F(4, 269) = 7.81$, $p < .0001$, and we found that quiz number significantly predicted quiz score overall calibration ($\beta_1 = .18$, $p = .001$) as did trait anxiety ($\beta_4 = .15$, $p = .0004$), but pre-quiz state anxiety ($\beta_2 = .06$, $p = .28$) and post-quiz state anxiety ($\beta_3 = -.02$, $p = .76$) did not (Table 2). The interrupted regression model took the same form as the multiple regression model and revealed no significant U-shaped relationship ($\beta_{1-5} = .20$, $p = .007$; $\beta_6 = -.08$, $p = .80$). These findings indicate a significant positive linear relationship for quiz score overall calibration through time, as when controlling for both state and trait anxiety, the overall quiz score calibration for each student increased with each subsequent quiz throughout the quarter (i.e., students became more underconfident over time); this finding is consistent with prior work showing that students may become more underconfident as their performance improves with practice (Koriat et al., 2002).

Finally, to examine how the difference in calibration from pre-quiz to post-quiz changed throughout time, we computed calibration difference scores (the absolute value of the difference between pre-quiz and post-quiz calibration) each week, for each student in the course. We then conducted a 1 (Exam: *quiz*) × 6 (Quiz number: *1, 2, ..., 6*) repeated-measures ANOVA on students' calibration difference scores, which revealed a significant main effect of quiz number, $F(3.62, 170.26) = 4.12$, $p = .004$, $\eta^2 = .08$ (Table 1). Post hoc paired-samples *t*-tests with a Bonferroni correction indicated significantly lower calibration differences scores for weeks 3 ($M = 1.05$, $SD = 1.37$), 4 ($M = 1.01$, $SD = 1.57$), 5 ($M = 1.01$, $SD = 1.22$), and 6 ($M = .91$, $SD = 1.32$) compared to week 2 ($M = 1.95$, $SD = 2.02$), adjusted *p*s < .02. No other comparisons were significant, adjusted *p*s > .74. Next, we used a multiple regression model that took the following form: calibration difference score = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (pre-quiz state anxiety) + $\beta_3$ (post-quiz state anxiety) + $\beta_4$ (trait anxiety). The four predictors significantly explained 7.1% of the variance, $R^2 = .08$, $F(4, 269) = 6.25$, $p < .0001$, and we found that quiz number significantly predicted quiz calibration difference scores ($\beta_1 = -.13$, $p = .004$), as did pre-quiz state anxiety ($\beta_2 = .13$, $p = .003$) and post-quiz state anxiety ($\beta_3 = -.12$, $p = .002$), but trait anxiety ($\beta_4 = .04$, $p = .17$) did not (Table 2). The interrupted regression model took the same form as the multiple regression model and revealed no significant U-shaped relationship ($\beta_{1–5} = -.15$, $p = .01$; $\beta_6 = -.24$, $p = .38$). Thus, we revealed that metacognitive calibration scores appear to be more discrepant from pre- to post-quiz earlier on in the quarter, becoming more stable from quiz 3 onward (see Fig. 3c).

## Anxiety Throughout the Quarter

To examine pre-quiz state anxiety with respect to quiz number, we conducted a 1 (Exam: *quiz*) × 6 (Quiz number: *1, 2, ..., 6*) repeated-measures ANOVA on pre-quiz state anxiety, which revealed a significant main effect of quiz number, $F(2.21, 10.26) = 4.66$, $p < .001$, $\eta^2 = .11$ (Table 1). Post hoc paired-samples *t*-tests with a Bonferroni correction indicated significantly lower pre-quiz state anxiety for quizzes 3 ($M = 5.73$, $SD = 2.33$) and 4 ($M = 5.49$, $SD = 2.20$) relative to quiz 1 ($M = 6.93$, $SD = 2.06$), adjusted *p*s < .006. No other comparisons were significant, adjusted *p*s > .31. Next, we conducted separate regressions examining pre-quiz and post-quiz state anxiety, each using both quiz number and trait anxiety as predictors. For pre-quiz state anxiety, the multiple regression model took the following form: pre-quiz state anxiety = $\beta_0$ + β1 (quiz number) + $\beta_2$ (trait anxiety). Two predictors significantly explained 17.0% of the variance, $R^2 = .18$, $F(2, 277) = 29.60$, $p < .0001$, and we found that trait anxiety significantly predicted pre-quiz state anxiety ($\beta_2 = .37$, $p < .0001$), but quiz number ($\beta_1 = -.13$, $p = .06$) did not (Table 3). The interrupted regression model took the same form as the multiple regression model and revealed no significant U-shaped relationship ($\beta_{1–3} = -.56$, $p = .009$; $\beta_{4–6} = .18$, $p = .20$). Thus, although there was a significant main effect of quiz number, controlling for trait anxiety revealed there to be no significant linear relationship between quiz number and pre-quiz state anxiety (see Fig. 4a).

Secondly, to examine post-quiz state anxiety with respect to quiz number, we conducted a 1 (Exam: *quiz*) × 6 (Quiz number: *1, 2, ..., 6*) repeated-measures ANOVA on post-quiz state anxiety, which revealed a significant main effect of quiz number, $F(5, 170) = 3.16$, $p = .009$, $\eta^2 = .09$ (Table 1). Post hoc paired-samples *t*-tests with a Bonferroni correction indicated significantly lower post-quiz state anxiety for quizzes 2 ($M = 4.76$, $SD = 2.44$) and 4 ($M = 4.74$, $SD = 2.28$) relative to quiz 1 ($M = 5.99$, $SD = 2.21$), adjusted *p*s < .019. No other comparisons were significant, adjusted *p*s > .056. For post-quiz state anxiety, the multiple regression model took the following form: post-quiz state anxiety = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (trait anxiety). Two predictors significantly explained 8.8% of the variance, $R^2 = .09$, $F(2, 271) = 14.14$, $p < .0001$, and we found that trait anxiety significantly predicted pre-quiz state anxiety ($\beta_2 = .27$, $p < .0001$), but quiz number ($\beta_1 = -.12$, $p = .12$) did not (Table 3). The interrupted regression model took the same form as the multiple regression model and revealed no significant U-shaped relationship ($\beta_{1–3} = -.65$, $p = .004$; $\beta_{4–6} = .09$, $p = .50$). Therefore,

**Table 3** Multiple regression analyses coefficients for state anxiety in study 1

| Outcome | % variance explained | $R^2$ | $F$ | $df$ | Intercept | $\beta_1$ | $\beta_2$ |
|---|---|---|---|---|---|---|---|
| Pre-quiz state anxiety | 17.0% | .18**** | 29.60 | (2, 277) | 4.37**** | −.13 | .37**** |
| Post-quiz state anxiety | 8.8% | .09**** | 14.14 | (2, 271) | 3.80**** | −.12 | .27**** |
| Overall state anxiety | 15.8% | .16**** | 27.09 | (2, 277) | 4.12**** | −.14* | .32**** |
| State anxiety difference score | 1.0% | .02 | 2.31 | (2, 271) | 1.78*** | −.09 | .05 |

*$p < .05$, **$p < .01$, ***$p < .001$, ****$p < .0001$

Model: Outcome = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (trait anxiety)

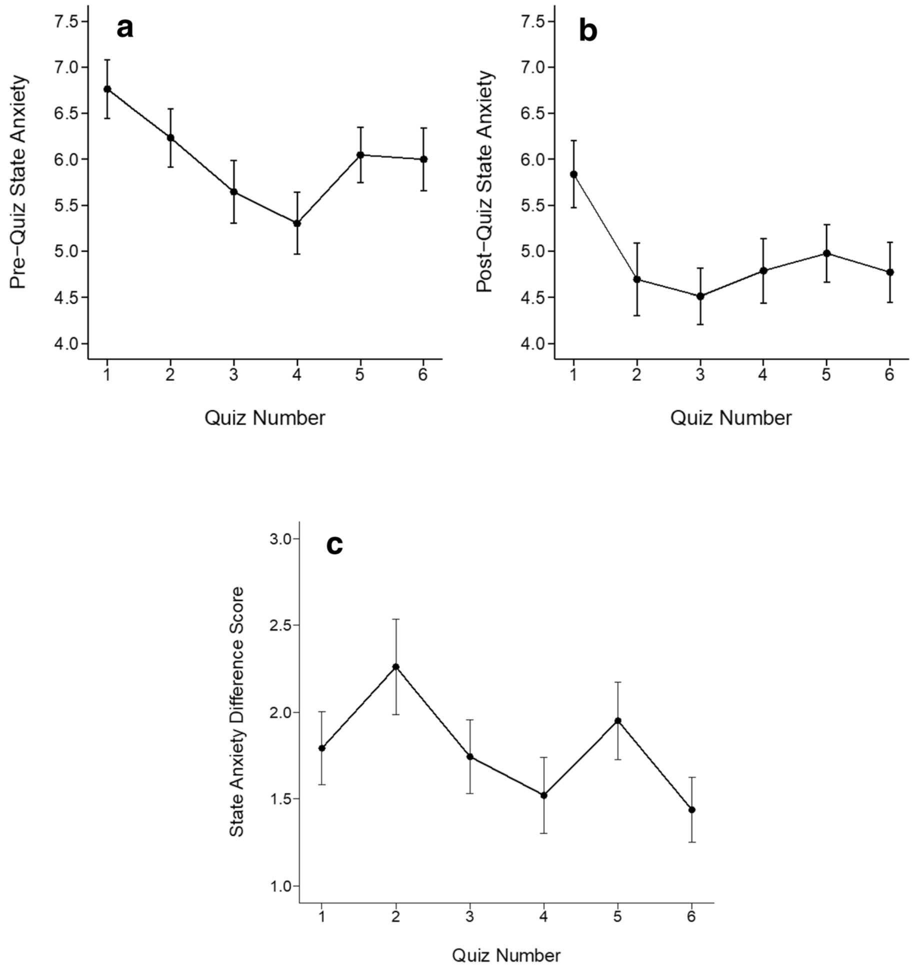


**Fig. 4** Mean state anxiety measures before (pre-quiz; panel **a**) and after (post-quiz; panel **b**) each quiz in study 1 across the academic quarter. State anxiety difference score (*absolute value of difference between pre-quiz and post-quiz anxiety*; panel **c**). All error bars represent ± 1 standard error

even though there was a main effect of quiz number, controlling for trait anxiety revealed no significant relationship between quiz number and post-quiz state anxiety (see Fig. 4b).

Thirdly, to examine overall state anxiety with respect to quiz number controlling for trait anxiety, we used a multiple regression model of the form: overall state anxiety = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (trait anxiety). Two predictors significantly explained 15.8% of the variance, $R^2 = .16$, $F(2, 277) = 27.09$, $p < .0001$, and we found that quiz number significantly predicted overall state anxiety ($\beta_1 = -.14$, $p = .04$) as did trait anxiety ($\beta_2 = .32$, $p < .0001$) (Table 3). The interrupted regression model took the same form as the multiple regression model and revealed no significant U-shaped relationship ($\beta_{1–5} = .91$, $p < .0001$; $\beta_{6–10} = 0$, $p = .99$). Thus, we revealed that overall state anxiety appears to decrease throughout the quarter even when controlling for trait anxiety.

Lastly, to examine how the difference in state anxiety changed from pre-quiz to post-quiz throughout time, we computed state anxiety difference scores (the absolute value of the difference between pre-quiz and post-quiz state anxiety) each week, for each student in the course. We then conducted a 1 (Exam: *quiz*) × 6 (Quiz number: *1, 2, ..., 6*) repeated-measures ANOVA on students' state anxiety difference scores, which revealed a significant main effect of quiz number, $F(5, 235) = 3.81$, $p = .002$, $\eta_p^2 = .08$ (Table 1). Post hoc paired-samples *t*-tests with a Bonferroni correction indicated significantly lower state anxiety differences scores for weeks 4 ($M = 1.53$, $SD = 1.49$) and 6 ($M = 1.44$, $SD = 1.30$) compared to week 2 ($M = 2.55$, $SD = 2.28$), adjusted *p*s < .006. No other comparisons were significant, adjusted *p*s > .057. Next, we used a multiple regression model that took the following form: anxiety difference score = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (trait anxiety). The two predictors did not significantly predict state anxiety difference scores, $R^2 = .02$, $F(2, 271) = 2.31$, $p = .10$; $\beta_1 = -.09$, $p = .12$; $\beta_2 = .05$, $p = .16$ (Table 3). The interrupted regression model took the same form as the multiple regression model and revealed no significant U-shaped relationship ($\beta_{1\text{–}2} = .47$, $p = .18$; $\beta_{3\text{–}6} = -.15$, $p = .047$). Thus, we revealed that state anxiety difference scores tend to vary with quiz number as indicated by the significant main effect (see Fig. 4c).

### Anxiety from Pre-quiz to Post-quiz

To test our hypothesis that state anxiety would decrease from pre- to post-quiz, we conducted a paired samples *t*-test comparing pre-quiz anxiety and post-quiz anxiety averaged across all six quizzes. The *t*-test was significant, confirming that pre-quiz anxiety ($M = 5.97$, $SD = 1.82$) was significantly higher than post-quiz anxiety ($M = 4.92$, $SD = 1.78$); $t(47) = 5.39$, $p < .001$, Cohen's $d = .78$.

### Anxiety and Quiz Performance

We first computed Pearson's correlations between overall quiz performance and measures of trait anxiety, overall state anxiety, pre-quiz anxiety, and post-quiz anxiety. The correlations between quiz performance and trait anxiety, $r(46) = -.008$, $p = .96$, quiz performance and overall state anxiety, $r(46) = -.28$, $p = .05$, and quiz performance and pre-quiz anxiety, $r(46) = -.10$, $p = .49$, were not significant. The correlation between quiz performance and post-quiz anxiety was significant, $r(46) = -.42$, $p = .003$, thus indicating that students' quiz performance varied inversely with their anxiety after taking each quiz.

Due to the possibility that Pearson's correlations between overall quiz performance and measures of state anxiety could be inflated due to the overlap of their variances, we computed partial Pearson's correlations that held trait anxiety constant. The partial correlations between quiz performance and overall state anxiety, $r(45) = -.32$, $p = .03$, and post-quiz anxiety, $r(45) = -.46$, $p = .001$, were significant, while the partial correlation between quiz performance and pre-quiz anxiety, $r(45) = -.11$, $p = .45$, was not significant. These analyses suggest, rather, that when holding trait anxiety constant, students' quiz performance varied inversely with both their overall state anxiety and post-quiz anxiety.

### Anxiety and Estimation Accuracy

We first computed Pearson's correlations between overall score calibrations and measures of trait anxiety, overall state anxiety, pre-quiz anxiety, and post-quiz anxiety. The correlation between overall score calibrations and trait anxiety was significant, $r(46) = .36$, $p = .01$, indicating that metacognitive accuracy varied with levels of trait anxiety; however, the correlations between overall score calibrations and overall state anxiety, $r(46) = .16$, $p = .28$, pre-quiz anxiety, $r(46) = .19$, $p = .19$, and post-quiz anxiety, $r(46) = .10$, $p = .51$, were not significant.

Due to the possibility that Pearson's correlations between overall score calibrations and measures of state anxiety could be inflated due to the overlap of their variances, we computed partial Pearson's correlations that held trait anxiety constant. None of the partial correlations between overall score calibrations and overall state anxiety, $r(45) = -.02$, $p = .87$, pre-quiz anxiety, $r(45) = .01$, $p = .93$, and post-quiz anxiety, $r(45) = -.06$, $p = .71$, were significant. Thus, when holding trait anxiety constant, we no longer find significant relationships between any measure of anxiety and estimation accuracy.

Lastly, we examined the relationship between state anxiety and metacognition and how they might interact to influence overall quiz performance. Table 3 illustrates the different models that were tested. We found there to be a significant moderation effect such that the effect of post-quiz anxiety on quiz performance seems to significantly depend on participants' metacognitive prediction calibration, $R^2 = .39$, $F(4, 269) = 43.40$, $p < .0001$; $\beta_{int} = .03$, $p = .01$.

We also found there to be a significant moderation effect such that the effect of post-quiz anxiety on quiz performance seems to significantly depend on participants' metacognitive postdiction calibration, $R^2 = .32$, $F(4, 269) = 31.71$, $p < .0001$; $\beta_{int} = -.05$, $p = .01$. Consequently, there was also a significant moderation effect of post-quiz anxiety on quiz performance such that it significantly depends on the metacognitive calibration difference score, $R^2 = .18$, $F(4, 269) = 14.75$, $p < .0001$; $\beta_{int} = -.05$, $p = .02$. No other interactions were significant, all $p$s > .06 (see Table 3).

### Study 1 Results Summary

In study 1, quiz performance increased with each subsequent quiz and post-quiz state anxiety was a predictor of quiz performance, while pre-quiz state anxiety and trait anxiety were not related to performance. Metacognitive pre- and postdictions did not vary with quiz when controlling for state and trait anxiety. Metacognitive pre- and postdiction calibration increased across the quarter when controlling for state and trait anxiety, and post-quiz anxiety was also a predictor of postdiction calibration. Thus, while students' performance increased throughout the quarter, they became increasingly more underconfident with each subsequent quiz indicating that they were underestimating their performance. Overall state anxiety decreased with quiz number, but this varied with trait anxiety, such that students with higher trait anxiety had higher overall state anxiety. Finally, the effect of post-state anxiety on performance was moderated by pre- and postdiction calibrations and pre-postdiction difference scores (Table 4).

## Study 2

### Method

### Participants

Participants consisted of 299 UCLA undergraduate students enrolled in Psychology 120A: Cognitive Psychology during Spring Quarter 2018 (92 students were excluded for missing data and 4 students were excluded as outliers: 203 students included in the final sample. Excluded outliers ±2.5 SDs from $Z$-score for overall score calibration).

## Materials and Procedure

There were two, non-cumulative exams administered during weeks 5 and 10. Exams were 60 multiple-choice questions based on the previous four weeks' worth of information. Scores ranged from 0 to 60, and exams accounted for 80% of the final grade (each exam worth 40%). An example of the type of multiple-choice question a student may have answered on an exam would be of the following form and level of conceptual understanding: *"Which of the following tasks is least appropriate as a means of testing implicit memory?"* There was also a written assignment, research proposal, and in-class presentation that factored into the final grade. The procedure for each exam was the same as the procedure for each quiz administered in study 1 and the same single-report scales were used (pre-quiz anxiety measure, prediction, test, post-anxiety measure, postdiction). The only difference was that trait anxiety was measured after each of the two exams. Students were informed that their participation in the surveys would not affect their test grades.

Similar to study 1, assigned course teaching assistants graded all assignments for the course and were blind to the hypotheses and predictions of the current study, as well as students' metacognitive ratings and reported anxiety measures. Additionally, all materials and procedures were performed ethically and approved by the UCLA Institutional Review Board.

**Table 4** How metacognition moderates effects of state anxiety on performance in study 1

| Model | % variance explained | $R^2$ | $F$ | $df$ | Intercept | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_{int}$ |
|---|---|---|---|---|---|---|---|---|---|
| A | 33.3% | .34**** | 35.80 | (4, 275) | 8.69**** | .14*** | .26* | −.18*** | .02 |
| B | 29.1% | .30**** | 29.66 | (4, 275) | 8.32**** | .15*** | .23* | −.11** | .02 |
| C | 37.7% | .39**** | 42.36 | (4, 269) | 8.63**** | .13*** | .35** | −.18**** | .01 |
| D | 36.4% | .37**** | 40.98 | (4, 275) | 8.60**** | .15*** | .16 | −.17**** | .03 |
| E | 33.3% | .34**** | 35.75 | (4, 275) | 8.27**** | .15**** | .24** | −.11*** | .02 |
| F | 38.3% | .39**** | 43.40 | (4, 269) | 8.57**** | .14*** | .14 | −.18**** | **.03*** |
| G | 21.1% | .22**** | 19.64 | (4, 275) | 8.69**** | .17**** | .29** | −.15**** | −.01 |
| H | 16.2% | .17**** | 14.43 | (4, 275) | 8.19**** | .18**** | .22* | −.06 | −.004 |
| I | 31.0% | .32**** | 31.71 | (4, 269) | 8.86**** | .12** | .60**** | −.19**** | **−.05*** |
| J | 12.1% | .13**** | 10.59 | (4, 275) | 8.44**** | .19**** | .12 | −.08 | −.03 |
| K | 8.48% | .10**** | 7.46 | (4, 275) | 8.10**** | .20**** | .04 | −.02 | −.02 |
| L | 16.8% | .18**** | 14.75 | (4, 269) | 8.49**** | .19**** | .14 | −.09* | **−.05*** |

*$p < .05$, **$p < .01$, ***$p < .001$, ****$p < .0001$

Significant interactions are **bolded**. Model A: Quiz score = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (overall calibration) + $\beta_3$ (overall state anxiety) + $\beta_{int}$ ($\beta_0$ * $\beta_3$). Model B: Quiz score = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (overall calibration) + $\beta_3$ (pre-quiz state anxiety) + $\beta_{int}$ ($\beta_2$ * $\beta_3$). Model C: Quiz score = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (overall calibration) + $\beta_3$ (post-quiz state anxiety) + $\beta_{int}$ ($\beta_2$ * $\beta_3$). Model D: Quiz score = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (pre-quiz calibration) + $\beta_3$ (overall state anxiety) + $\beta_{int}$ ($\beta_2$ * $\beta_3$). Model E: Quiz score = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (pre-quiz calibration) + $\beta_3$ (pre-quiz state anxiety) + $\beta_{int}$ $\beta_2$ * $\beta_3$). Model F: Quiz score = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (pre-quiz calibration) + $\beta_3$ (post-quiz state anxiety) + $\beta_{int}$ ($\beta_2$ * $\beta_3$). Model G: Quiz score = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (post-quiz calibration) + $\beta_3$ (overall state anxiety) + $\beta_{int}$ ($\beta_2$ * $\beta_3$). Model H: Quiz score = $\beta_0$ + $\beta_1$ (quiz number) + $\beta_2$ (post-quiz calibration) + $\beta_3$ (pre-quiz state anxiety) + $\beta_{int}$ ($\beta_2$ * $\beta_3$). Model I: Quiz score = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (post-quiz calibration) + $\beta_3$ (post-quiz state anxiety) + $\beta_{int}$ ($\beta_2$ * $\beta_3$). Model J: Quiz score = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (calibration difference score) + $\beta_3$ (overall state anxiety) + $\beta_{int}$ ($\beta_2$ * $\beta_3$). Model K: Quiz score = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (calibration difference score) + $\beta_3$ (pre-quiz state anxiety) + $\beta_{int}$ $\beta_2$ * $\beta_3$). Model L: Quiz score = $\beta_0 + \beta_1$ (quiz number) + $\beta_2$ (calibration difference score) + $\beta_3$ (post-quiz state anxiety) + $\beta_{int}$ ($\beta_2$ * $\beta_3$)

## Results and Discussion

### Exam Performance

In order to determine if students performed differently between exam 1 and exam 2, we conducted a paired samples *t*-test between students' scores for each exam. Scores improved from exam 1 ($M$ = 49.67, $SD$ = 5.91) to exam 2 ($M$ = 51.90, $SD$ = 4.92), $t(206) = 6.67$, $p < .001$, Cohen's $d$ = .46.

### Estimation Accuracy

We conducted a 2 (Exam: *midterm, final*) × 2 (Calibration: *pre-exam, post-exam*) repeated-measures ANOVA on overall score calibration to examine how well students metacognitively estimated their scores before and after each exam (see Fig. 5). There was a significant main effect of exam such that overall score calibration was significantly more positive (i.e., more underconfident) for exam 2 ($M$ = 2.13, $SD$ = 5.61) compared to exam 1 ($M$ = .33, $SD$ = 5.97), regardless of whether these metacognitive estimations of accuracy were provided before or after each exam, $F(1, 202) = 13.56$, $p < .001$, $\eta^2 = .04$. These analyses also revealed a

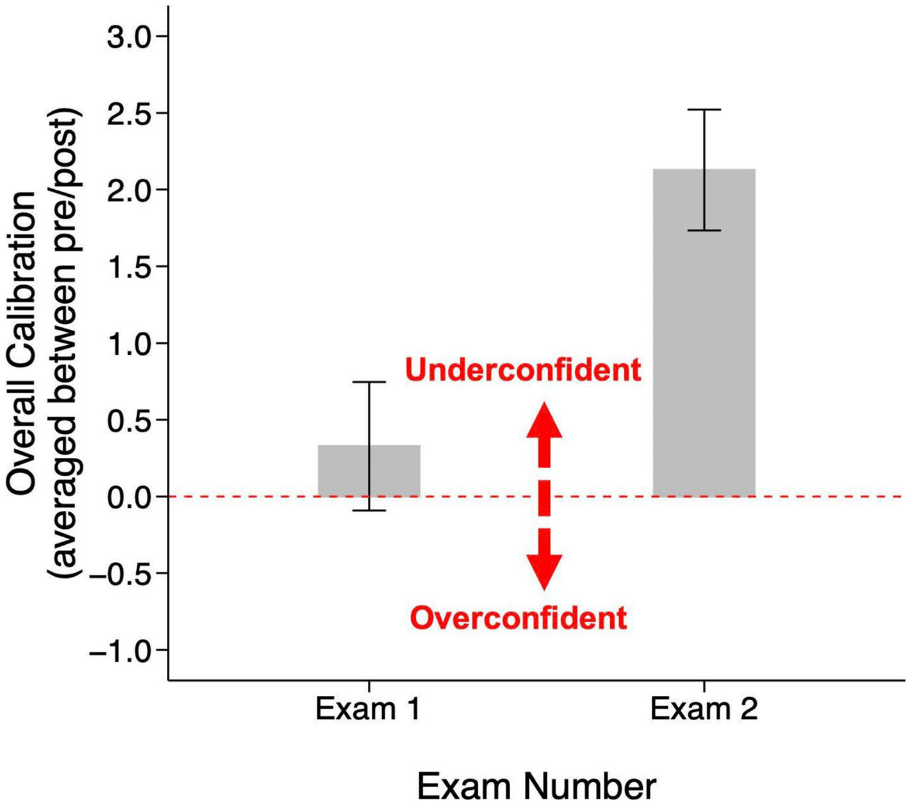


**Fig. 5** Mean overall metacognitive score calibration (*actual score − predicted/postdicted score*) measures for both exams in study 2. More positive scores indicate higher underconfidence. All error bars represent ± 1 standard error

significant main effect of calibration such that postdiction score calibration ($M$ = 1.83, $SD$ = 4.59) was significantly more underconfident than was prediction score calibration ($M$ = .63, $SD$ = 5.32), $F(1, 202) = 22.42$, $p < .001$, $\eta^2 = .02$. Finally, there was a significant interaction between exam and calibration, $F(1, 202) = 9.60$, $p = .002$, $\eta^2 = .005$. Follow-up post hoc independent samples $t$-tests with a Bonferroni correction revealed there to be significant differences between prediction calibration for exam 1 ($M$ = .18, $SD$ = 7.64) and exam 2 ($M$ = − 2.20, $SD$ = 7.58), between pre- ($M$ = .18, $SD$ = 7.64) and postdiction ($M$ = −1.69, $SD$ = 7.32) calibration for exam 1, and between the prediction calibration for exam 1 ($M$ = .18, $SD$ = 7.64) and the postdiction calibration for exam 2 ($M$ = −3.00, $SD$ = 6.84) (adjusted $p$s < .001).

## Anxiety Across Exams

In order to determine whether or not measures of anxiety remained relatively constant from exam 1 to exam 2, we conducted multiple paired samples $t$-tests between trait anxiety, pre-exam anxiety, and post-exam anxiety for exam 1 and exam 2. We found no significant differences for any measure of anxiety between exam 1 and exam 2, indicating that students were no more anxious for the final exam compared to the midterm exam: trait anxiety at exam 1 ($M$ = 6.22, $SD$ = 2.36) vs. exam 2 ($M$ = 6.37, $SD$ = 2.34), $t(206) = 1.66$, $p = .10$, Cohen's $d$ = .12; pre-exam anxiety for exam 1 ($M$ = 6.28, $SD$ = 2.18) vs. exam 2 ($M$ = 6.50, $SD$ = 2.39), $t(206) = 1.44$, $p = .15$, Cohen's $d$ = .10; post-exam anxiety for exam 1 ($M$ = 5.85, $SD$ = 2.39) vs. exam 2 ($M$ = 5.83, $SD$ = 2.44), $t(206) = .12$, $p = .91$, Cohen's $d$ = .01.

## Anxiety from Pre- to Post-exam

To test our hypothesis that anxiety would decrease from pre- to post-exam, we conducted a paired samples $t$-test comparing pre-exam anxiety and post-exam anxiety averaged across both exams. The $t$-test was significant confirming that pre-exam anxiety ($M$ = 6.35, $SD$ = 2.00) was significantly higher than post-exam anxiety ($M$ = 5.80, $SD$ = 2.18); $t(202) = 4.50$, $p < .001$, Cohen's $d$ = .32.

## Anxiety and Exam Performance

Similar to study 1, we first computed Pearson's correlations between overall exam performance and measures of trait anxiety, overall state anxiety, pre-exam anxiety, and post-exam anxiety. The correlations between exam performance and trait anxiety, $r(201) = −.20$, $p$ = .004, overall state anxiety, $r(201) = −.32$, $p < .001$, pre-exam anxiety, $r(201) = −.25$, $p < .001$, and post-exam anxiety, $r(201) = −.33$, $p < .001$, were all significant, thus indicating that students' exam performance varied inversely with all measures of their anxiety before and after taking each exam.

Then, due to the possibility that Pearson's correlations between overall exam performance and measures of state anxiety could be inflated due to the overlap of their variances, we computed partial Pearson's correlations holding trait anxiety constant. All partial correlations between overall exam performance and overall state anxiety, $r(200) = −.26$, $p < .001$, pre-exam anxiety, $r(200) = −.17$, $p = .01$, and post-exam anxiety, $r(200) = −.27$, $p < .0001$, were significant, thus indicating that students' exam performance varied inversely with all measures of their state anxiety before and after taking each exam when holding trait anxiety constant.

### Anxiety and Estimation Accuracy

We also computed Pearson's correlations between overall score calibrations and measures of trait anxiety, overall state anxiety, pre-exam anxiety, and post-exam anxiety. None of the correlations between overall score calibrations and trait anxiety, $r(201) = .02$, $p = .73$, overall state anxiety, $r(201) = .07$, $p = .32$, pre-exam anxiety, $r(201) = .09$, $p = .19$, and post- exam anxiety, $r(201) = .04$, $p = .59$, were significant.

Then, due to the possibility that Pearson's correlations between overall score calibrations and measures of state anxiety could be impacted by trait anxiety, we computed partial Pearson's correlations holding trait anxiety constant, and found that none of the partial correlations between overall score calibrations and overall state anxiety, $r(200) = .07$, $p = .33$, pre-exam anxiety, $r(200) = .09$, $p = .18$, and post-exam anxiety, $r(200) = .03$, $p = .67$, were significant. Thus, when holding trait anxiety constant, we still do not find significant relationships between any measure of anxiety and estimation accuracy.

## Study 2 Results Summary

In study 2, higher state anxiety (pre- and post-quiz and overall) was associated with lower exam performance when controlling for trait anxiety. Neither state nor trait anxiety impacted the estimation accuracy of performance for participants. Students also became more underconfident from exam 1 to exam 2 while their scores increased, indicating that students were metacognitively unaware of their performance gains.

## General Discussion

The goal of the current research was to explore the effects of test anxiety on undergraduate students' overall performance and metacognitive accuracy and how the frequency and stakes of the assessment influence those effects. In both studies, we found that state anxiety decreased from before to after assessments, and overall score calibration increased throughout the academic quarter (see Table 5 for a summary of the convergence and divergence of results from study 1 to study 2). Additionally, higher post-exam state anxiety was associated with worse assessment performance. Although students were less anxious once they completed the quiz or exam, they did not recognize that their performance increased. Students were more anxious after the exam when they knew they had performed poorly (or were not anxious knowing that they had done well).

In study 1, we found there to be a main effect of quiz number on pre- and post-quiz state anxiety; however, regression analyses controlling for trait anxiety revealed there to be no significant linear relationship between quiz number and pre/post-quiz state anxiety. Overall, students appeared to be more anxious at the beginning of the quarter relative to the middle of the quarter, with state anxiety generally stabilizing after acclimation to the course. Accordingly, since pre- and post-state anxiety did not decrease with quiz in study 1, this further suggests that the impact of frequent, low-stakes quizzes on state anxiety may vary with individual differences in trait anxiety. Similarly, pre-exam state anxiety was constant across the two exams (middle and end of the term) in study 2.

**Table 5** Comparison of the main findings across studies

| Results summary | Study 1 | Study 2 |
|---|---|---|
| Assessment performance[1] | + | + |
| Overall calibration[1] | + | + |
| Pre-assessment state anxiety[1] | – | No change |
| Post-assessment state anxiety[1] | – | No change |
| Changes in state anxiety[2] | – | – |
| Overall state anxiety and performance[3] | – | – |

"+" indicates an increase across time points or a positive correlation while "–" indicates a decrease across time points or a negative correlation

[1] Throughout the quarter

[2] From pre-assessment to post-assessment

[3] Partial Pearson's correlation controlling for trait anxiety

In study 1, trait anxiety varied positively with overall score calibration, but was not associated with quiz performance; however, in study 2, higher trait anxiety was associated with poorer exam performance, but not with estimation accuracy. Specifically, we found that students' metacognitive predictions increased throughout the term in both studies, in addition to overall performance increasing throughout the term for both classes. Additionally, in study 1, overall score calibrations increased with each subsequent quiz when controlling for both state and trait anxiety, which illustrates that students became more underconfident with practice even though their performance increased across the quarter. Similarly, students became more underconfident from exam 1 to exam 2 in study 2, despite their increase in performance.

Metacognitive postdictions increased throughout the term in study 1 but remained unchanged in study 2. Though, this increase in pre-/postdictions in study 1 can be accounted for by holding trait anxiety constant in analyses. When considering individual differences in trait anxiety, pre-postdictions stayed constant across the quarter suggesting that metacognitive predictions are related to trait anxiety. Additionally, prediction calibration increased from exam 1 to exam 2 in study 2, and both prediction calibration and postdiction calibration increased throughout the term in study 1; trait anxiety was a predictor of this increase, but state anxiety was not related to this increase. This is consistent with our findings that trait anxiety seems to be related to metacognition when taking frequent, low-stakes quizzes. Though, postdiction calibration did not significantly change throughout the term in study 2 while overall score calibration did increase throughout the term for both courses. As such, this observed change in metacognitive performance throughout the term in both studies reveals a potential "underconfidence with practice effect" (Koriat et al., 2002), such that students became more underconfident throughout the quarter upon completing more assessments.

Students in study 1 with higher trait anxiety exhibited higher metacognitive underconfidence when taking the more frequent, lower-stakes quizzes, though this trait anxiety was not related to the students' overall performance. Yet, when controlling for trait anxiety, there does not appear to be a significant relationship between state anxiety and metacognitive accuracy. Conversely, trait anxiety was not related to students' metacognitive accuracy in study 2, despite its relationship with overall performance. These results reveal a critical implication for determining testing practices that are more inclusive to individuals with higher levels of baseline anxiety. The current findings are compelling in that they reveal the importance of considering students' levels of test anxiety to motivate the implementation of more frequent, lower-stakes quizzes in courses. A primary goal for educators and those designing more modular and adaptable course curricula at all levels of education should be to consider the various intrinsic (Deci et al., 1999) and extrinsic (Ryan & Weinstein, 2009) factors that may perpetually enhance (Putwain et al., 2012; Vansteenkiste et al., 2004) or decline (Ratelle et al., 2007) the students' educational outcomes in the classroom.

To further examine how metacognitive accuracy influences the relationship between state anxiety and performance, we ran a series of regression models, which revealed that metacognitive accuracy was a moderator of the effects of post-quiz state anxiety on performance in study 1. Specifically, pre- and postdiction calibrations and pre-postdiction calibration difference scores moderated the relationship between post-quiz state anxiety on performance (see Table 4).

Thus, it seems that although we found that higher post-quiz state anxiety is related to lower performance outcomes, this varies with prediction accuracy, postdiction accuracy, and differences between pre- and postdiction measures. Individual differences in metacognitive abilities seem to be related to the effects of post-state anxiety on performance when taking more frequent, lower-stakes quizzes.

We should also note that although many of our regression models were significant in study 1, $R^2$ values were relatively small, explaining between 3 and 17% of the variance amongst the data. The implications of these results indicate an inherent amount of unexplainable variability within our dataset, most likely due to the practical considerations that our sample was constrained to a real, classroom setting comprised of enrolled students in the course, as opposed to a controlled, randomly sampled in-lab group of recruited participants.

More broadly, integrating optimized tools and techniques to develop more equitable and customized approaches for different students to achieve learning outcomes, regardless of their background and preparation, is a critical consideration that must be accounted for in future course planning. For instance, students from lower socioeconomic backgrounds may already be prone to higher levels of stress and anxiety (Lantz et al., 2005), which may further instigate higher levels of test anxiety in these individuals. Additionally, there are intergenerational effects of certain academic anxieties (Maloney et al., 2015) such that students of parents with math anxiety experience lower math achievement when their parents have been highly involved in their math development. These considerations are essential when designing assessments that provide optimal and equitable conditions for learning achievement. In fact, it is vital to consider students' diverse backgrounds at all stages of the test cycle as learners with high test anxiety have been shown to have deficits in encoding, organization, and storage of learned information (Mueller, 1980). These deficits during the test preparation phase often lead to poor test performance for students with high test anxiety such that anxiety prompts interference, distractibility, or inefficient cue utilization strategies (Cassady, 2004; Schwarzer & Jerusalem, 1992). During the test reflection phase, learners typically form attributional biases, which can lead to negative future attitudes and behavior patterns towards testing. If a student attributes their failure on the assessment to a lack of ability, they are more likely to avoid proper test preparation activities for future tests (Elliot & McGregor, 1999).

Some limitations of the current study include only investigating student anxiety both immediately before and after the assessment without considering other factors that could potentially explain why these effects were observed. More specifically, we did not consider how the differences in low-stakes quizzes may prompt students to adjust their study practices throughout the term (study 1) compared to a different class with a completely different assessment structure (study 2). Therefore, future research should examine how students' individual levels of preparation for various types of assessments mediate their levels of post-quiz anxiety and performance. Additionally, future classroom-based studies examining test anxiety should seek to collect a variety of demographic factors, along with reports and measures of students' efforts and strategies to prepare for quizzes and exams, to subsequently compare differences in test anxiety both within and across groups of students at all stages of the testing cycle. These findings, when considered with students' individual circumstances, may provide a framework for designing more equitable assessments that take into account factors potentially interacting with or provoking students' test anxiety. It is also worth noting that although single-report measures were most practical for this study, future research should consider these relationships between characteristics of assessments (i.e., format, frequency, and stakes) and state and trait anxiety while using validated scales to measure individual differences in different types of test anxiety.

In conclusion, the findings of the current study suggest that less-frequent, higher-stakes assessments may trigger students' baseline levels of anxiety in that students may inherently feel more pressure to perform better on these exams compared to testing environments with more-frequent, lower-stakes quizzes, as each individual assessment would not necessarily have as substantial of an impact on a student's overall course grade comparatively. This finding complements prior work showing that students are overwhelmed with stress, anxiety, and worry due to testing in high-stakes contexts (e.g., Segool et al., 2013; Triplett & Barksdale, 2005).

Though, students' feelings regarding testing seem to be inconsistent across studies examining the relationship between test anxiety and the relative weight of an assessment (Mulvenon et al., 2001; Mulvenon et al., 2005; Putwain, 2008). Our findings, therefore, contribute to these prior equivocal results in the literature that prompted us to further investigate how the relationship between test anxiety and the stakes of an assessment vary with one another, albeit including an examination of how test anxiety interacts with students' metacognitive accuracy.

**Acknowledgements** We thank Gerardo Ramirez, Barbara Knowlton, Robert Bjork, Elizabeth Bjork, and Courtney Clark for helpful comments regarding this work. We also thank Mary Hargis, Dillon Murphy, Julia Schorn, Mary Whatley, and the rest of the members of the Memory and Lifespan Cognition Laboratory at UCLA for their guidance and support.

**Code Availability** No custom code or software were written for this project. Code to run analyses in R are provided in the OSF link below.

**Author Contribution** A. D. Castel and A. L. M. Siegel conceived of the study. A. L. M. Siegel collected the data, and S. T. Schwartz analyzed the data with A. L. M. Siegel and K. M. Silaj. K. M. Silaj and S. T. Schwartz wrote the manuscript. A. L. M. Siegel and A. D. Castel contributed to the editing of the manuscript.

**Funding** This research was supported in part by the National Institutes of Health (National Institute on Aging; Award Number R01 AG044335 to Dr. Alan Castel).

**Data Availability** Data and survey materials are available via OSF (https://osf.io/pv89n/).

## Declarations

**Conflict of Interest** The authors declare no competing interests.

## References

Argarwal, P. K., D'Antonio, L., Roediger, H. L. I., McDermott, K. B., & McDaniel, M. A. (2014). Classroom-based programs of retrieval practice reduce middle school and high school students' test anxiety. *Journal of Applied Research in Memory and Cognition, 3*(3), 131–139.

Birenbaum, M., & Feldman, R. A. (1998). Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research, 40*(1), 90–98.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: knowing about knowing* (pp. 185–205). MIT Press.

Bowman, A. W., Jones, M. C., & Gijbels, I. (1998). Testing monotonicity of regression. *Journal of Computational and Graphical Statistics, 7*(4), 489–500.

Callender, A. A., Franco-Watkins, A. M., & Roberts, A. S. (2016). Improving metacognition in the classroom through instruction, training, and feedback. *Metacognition and Learning, 11*(2), 215–235.

Cassady, J. C. (2004). The impact of test anxiety on text comprehension and recall in the absence of external evaluative pressure. *Applied Cognitive Psychology, 18*(3), 311–325.

Cassady, J. C. (2010). Test anxiety: contemporary theories and implications for learning. In J. C. Cassady (Ed.), *Anxiety in schools: the causes, consequences, and solutions for academic anxieties* (pp. 7–26). Peter Lang.

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary educational anxiety, 27*, 270–295.

Cizek, G. J., & Burg, S. S. (2006). *Addressing test anxiety in a high-stakes environment: strategies for classrooms and schools*. Corwin Press.

Covington, M. V. (1985). Test anxiety: causes and effects over time. In H. M. van der Ploeg, R. Schwarzer, & C. D. Spielberger (Eds.), *Advances in test anxiety research* (Vol. 4, pp. 55–68). Swets & Zeitlinger.

Davey, H. M., Barratt, A. L., Butow, P. N., & Deeks, J. J. (2007). A one-item question with a Likert or Visual Analog Scale adequately measured current anxiety. *Journal of Clinical Epidemiology, 60*(4), 356–360.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58.

Dunn, D. S., Saville, B. K., Baker, S. C., & Marek, P. (2013). Evidence-based teaching: tools and techniques that promote learning in the psychology classroom. *Australian. Journal of Psychology, 65*(1), 5–13.

Elliot, A. J., & McGregor, H. A. (1999). Test anxiety and the hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 76*(4), 628–644.

Everson, H. T., Smodlaka, I., & Tobias, S. (1994). Exploring the relationship of test anxiety and metacognition on reading test performance: a cognitive analysis. *Anxiety, stress, and coping, 7*(1), 85–96.

Fernández-Castillo, A., & Caurcel, M. J. (2015). State test-anxiety, selective attention and concentration in university students. *International Journal of Psychology, 50*(4), 265–271.

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do girls really experience more anxiety in mathematics? *Psychological Science, 24*(10), 2079–2087.

Harris, R. B., Grunspan, D. Z., Pelch, M. A., Fernandes, G., Ramirez, G., & Freeman, S. (2019). Can test anxiety interventions alleviate a gender gap in an undergraduate STEM course? *CBE—Life Sciences Education, 18*, 1–9.

Hembree, R. (1988). Correlates, causes, effect, and treatment of test anxiety. *Review of Educational Research, 58*(1), 47–77.

Karpicke, J. D. (2017). Retrieval-based learning: a decade of progress. In J. T. Wixted (Ed.), *Cognitive psychology of memory, of learning and memory: a comprehensive reference* (Vol. 2, pp. 487–514). Oxford: Academic Press.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131*(2), 147–162.

Kurosawa, K., & Harackiewicz, J. M. (1995). Test anxiety, self-awareness, and cognitive interference: a process analysis. *Journal of Personality, 63*(4), 931–951.

Lantz, P. M., House, J. S., Mero, R. P., & Williams, D. R. (2005). Stress, life events, and socioeconomic disparities in health: results from the Americans' Changing Lives Study. *Journal of Health and Social Behavior, 46*(3), 274–288.

Leadbeater, B., Thompson, K., & Gruppuso, V. (2012). Co-occurring trajectories of symptoms of anxiety, depression, and oppositional defiance from adolescence to young adulthood. *Journal of Clinical Child & Adolescent Psychology, 41*(6), 719–730.

Legg, A. M., & Locker, L. (2009). Math performance and its relationship to math anxiety and metacognition. *North American Journal of Psychology, 11*, 471–486.

Maloney, E. A., Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2015). Intergenerational effects of parents' math anxiety on children's math achievement and anxiety. *Psychological Science, 26*(9), 1480–1488.

Mueller, J. H. (1980). Test anxiety and the encoding and retrieval of information. In I. G. Sarason (Ed.), *Test anxiety: theory, research, and applications* (pp. 63–86). Erlbaum.

Mulvenon, S. W., Connors, J. V., & Lenares, D. (2001). Impact of accountability and school testing on students: is there evidence of anxiety?.

Mulvenon, S. W., Stegman, C. E., & Ritter, G. (2005). Test anxiety: a multifaceted study on the perceptions of teachers, principals, counselors, students, and parents. *International Journal of Testing, 5*(1), 37–61.

Naveh-Benjamin, M. (1991). A comparison of training programs intended for different types of test anxious students: further support for an information-processing model. *Journal of Educational Psychology, 83*(1), 134–139.

Núñez-Peña, M. I., Suárez-Pellicioni, M., Guilera, G., & Mercadé-Carranza, C. (2013). A Spanish version of the short Mathematics Anxiety Rating Scale (sMARS). *Learning and Individual Differences, 24*, 204–210.

Putwain, D. W. (2008). Deconstructing test anxiety. *Emotional and Behavioural Difficulties, 13*(2), 141–155.

Putwain, D. W., & Best, N. (2011). Fear appeals in the primary classroom: effects on test anxiety and test grade. *Learning and Individual Differences, 21*(5), 580–584.

Putwain, D. W., Kearsley, R., & Symes, W. (2012). Do creativity self-beliefs predict literacy achievement and motivation? *Learning and Individual Differences, 22*(3), 370–374.

Ratelle, C. F., Guay, F., Vallerand, R. J., Larose, S., & Senécal, C. (2007). Autonomous, controlled, and amotivated types of academic motivation: a person-oriented analysis. *Journal of Educational Psychology, 99*(4), 734–746.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255.

Ryan, R. M., & Weinstein, N. (2009). Undermining quality teaching and learning: a self- determination theory perspective on high-stakes testing. *Theory and Research in Education, 7*(2), 224–233.

Schwarzer, R., & Jerusalem, M. (1992). Advances in anxiety theory: a cognitive process approach. In K. A. Hagtvet & T. B. Johnsen (Eds.), *Advances in test anxiety research* (Vol. 7, pp. 2–31). Swets & Zeitlinger.

Segool, N. K., Carlson, J. S., Goforth, A. N., von der Embse, N., & Barterian, J. A. (2013). Heightened test anxiety among young children: elementary school students' anxious responses to high-stakes testing. *Psychology in the Schools, 50*(5), 489–499.

Simonsohn, U. (2018). Two lines: a valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science, 1*(4), 538–555.

Spada, M. M., Nikcevic, A. V., Moneta, G. B., & Ireson, J. (2006). Metacognition as a mediator of the effect of test anxiety on a surface approach to studying. *Educational Psychology, 26*(5), 615–624.

Spielberger, C. (1972). Anxiety as an emotional state. In C. D. Spielberger (Ed.), *Anxiety: current trends in theory and research* (Vol. 1, pp. 23–49). Academic Press.

Triplett, C. F., & Barksdale, M. A. (2005). Third through sixth graders' perceptions of high- stakes testing. *Journal of Literacy Research, 37*(2), 237–260.

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: the synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology, 87*(2), 246–260.

Veenman, M. V. J., Kerseboom, L., & Imthorn, C. (2000). Test anxiety and metacognitive skillfulness: availability versus production deficiencies. *Anxiety, Stress, and Coping, 13*(4), 391–412.

von der Embse, N., & Hasson, R. (2012). Test anxiety and high-stakes test performance between school settings: implications for educators. *Preventing school failure: alternative education for children and youth, 56*(3), 180–187.

von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: a 30-year meta-analytic review. *Journal of Affective Disorders, 227*, 483–493.

Williams, J. E. (1991). Modeling test anxiety, self-concept and high school students' academic achievement. *Journal of Research & Development in Education, 25*(1), 51–57.

Wood, S. G., Hart, S., Little, S., & Phillips, S. A. (2016). Test anxiety and a high-stakes standardized reading comprehension test: a behavioral genetics perspective. *Merrill-Palmer Quarterly, 62*(3), 233–251.

Yang, C., Sun, B., Potts, R., Yu, R., Luo, L., & Shanks, D. R. (2020). Do working memory capacity and test anxiety modulate the beneficial effects of testing on new learning? *Journal of Experimental Psychology: Applied* Advance online publication.

Yue, X. (1996). Test anxiety and self-efficacy: Levels and relationship among secondary school students in Hong Kong. *Psychologia: An International Journal of Psychology in the Orient, 39*(3), 193–202.

Zohar, D. (1998). An additive model of test anxiety: role of exam-specific expectations. *Journal of Educational Psychology, 90*(2), 330–340.