Miscalibration of self-efficacy and academic performance:

Self-efficacy ≠ self-fulfilling prophecy

Kate Talsma[a], Benjamin Schüz[b], Kimberley Norris[a]

a University of Tasmania, School of Medicine (Psychology), Tasmania, Australia

b University of Bremen, Human and Health Sciences, Bremen, Germany

Abstract

While it has been consistently demonstrated that academic self-efficacy and performance are positively correlated in groups of students, little is known about whether individual students' academic self-efficacy levels align with their own performance abilities. At the same time, researchers contest whether self-efficacy *should* align with performance abilities to be of most benefit to students. In this study, we applied procedures used in the meta-cognitive calibration paradigm to investigate the alignment between academic self-efficacy and academic performance (i.e., *self-efficacy calibration*) in higher education. Undergraduate students (n=207) completed five self-efficacy questionnaires with regard to academic performance outcomes in one subject over a semester (two written assignments, two exams, and the subject overall). Five corresponding grades were also collected. We calculated two types of self-efficacy calibration scores: self-efficacy accuracy (the deviation between self-efficacy and performance) and self-efficacy bias (the signed difference [i.e., valence]; over- and under-efficaciousness). Miscalibration of self-efficacy beliefs was prevalent, consistent with findings regarding meta-cognitive calibration. Under-efficaciousness was common at task level (for written assignments and exams), while over-efficaciousness was pronounced at domain level (for the subject overall). Self-efficacy exceeded performance for low-achievers, while it fell short of performance for high-achievers. A key finding was that self-efficacy bias predicted academic performance on similar subsequent tasks, with under-efficacious students performing better than accurate or over-efficacious students. Findings suggest self-efficacy is not a self-fulfilling prophecy; instead, over-efficacious students may experience negative impacts on academic self-regulation and performance.

Key words: Self-efficacy, academic performance, accuracy, bias, calibration

1

Self-efficacy – perceptions of one's capability to organise and execute required courses of action to achieve particular outcomes (Bandura, 1997) – is widely believed to be one of the most important non-intellective determinants of academic performance (e.g., Richardson, Abraham, & Bond, 2012; Schneider & Preckel, 2017). However, there is a growing understanding in the literature that the relationship between self-efficacy and academic performance is complex. For example, many factors moderate the relationship between the two, and reciprocity in the relationship is not evident under all circumstances (Honicke & Broadbent, 2016; Schöber, Schütte, Köller, McElvany, & Gebauer, 2018; Talsma, Schüz, Schwarzer, & Norris, 2018). These complexities may be partly explained by differences between individuals regarding the degree of match between subjective self-efficacy judgements and objective performance outcomes. The purpose of the present study was to determine how aligned university students' academic self-efficacy beliefs are with their academic performance outcomes, and to establish how alignment of self-efficacy and performance on one task predicts future performance outcomes on tasks of the same type.

The alignment of several types of subjective beliefs and corresponding objective outcomes has been explored within a calibration paradigm (e.g., Dunlosky & Thiede, 2013; Hacker, Bol, & Keener, 2008). For example, a substantial body of research has considered calibration of metacognitive judgements (see Schraw, 2009b for a review and taxonomy). Metacognitive judgements can include judgements made prior to performance (such as "judgements of learning" and "feeling of knowing" judgements), during performance (sometimes called "online" accuracy or confidence judgements), and after performance (which can include retrospective accuracy or confidence judgements). Calibration research has also been conducted investigating understanding of texts, known as calibration of comprehension (e.g., Stolp & Zabrucky, 2009), as well as the accuracy of test performance

predictions (e.g., Hacker, Bol, Horgan, & Rakow, 2000). On the whole, calibration research tends to find that people think they know more than they actually do (Zabrucky, 2010).

Most of the calibration research conducted thus far focuses on whether people *know what they know*. Calibration in this sense is important in educational settings, as it is believed to influence academic behaviours and outcomes (e.g., Dunlosky & Rawson, 2012; Ramdass & Zimmerman, 2008; Zabrucky, 2010). Calibration is argued to play a role in self-regulation, that is, "the self-directive process by which learners transform their mental abilities into academic skills" (Zimmerman, 2002, p. 65). Calibration indicates that accurate self-monitoring is occurring, which enables students to take the necessary actions to perform well (Bol & Hacker, 2012; Stone, 2000).

Very little research, however, specifically considers the calibration of beliefs which form part of a social-cognitive framework (Stolp & Zabrucky, 2009), such as self-efficacy: that is, whether people *can do what they believe they can do*. This question is future oriented and considers performance outcomes rather than knowledge specifically. Given the importance of self-efficacy-beliefs to effective academic self-regulatory capacity (Zabrucky, 2010; Zimmerman, 2002), an understanding of the calibration of academic self-efficacy is also critical. In Zimmerman's (2000, 2002) three-phase model of self-regulation, self-efficacy forms part of the pre-performance phase, where it contributes to a person's motivation in combination with other variables such as outcome expectations and task value. Self-regulation also includes a performance phase, in which the individual engages in self-monitoring and control processes, and a post-performance phase, in which the individual engages in attribution and reflection processes. Students' self-efficacy calibration may be influenced by monitoring processes during the performance phase, and by reflection processes in the post-performance phase.

Consistent significant positive correlations between self-efficacy and academic performance are often interpreted in the literature to mean that self-efficacy beliefs are accurate (i.e., calibrated) with respect to performance outcomes (e.g., Ackerman, Beier, & Bowen, 2002; Moores, Chang, & Smith, 2006). This provides support for the idea of self-efficacy as a self-fulfilling prophecy – believe, and you will achieve (Pajares, 2006). In educational settings, it is suggested that this self-fulfilling prophecy means that students with high self-efficacy persevere and perform well, while those with low self-efficacy give up and disengage (Pajares, 2002; Schunk & Pajares, 2004). However, the relationship between self-efficacy and academic performance is typically reported in meta-analyses to be moderate (Honicke & Broadbent, 2016; Richardson et al., 2012; Talsma et al., 2018), meaning that there is much variance still to be explained, and even a large positive correlation may be underscored by consistent bias (Dunning & Helzer, 2014). Furthermore, correlations between self-efficacy and academic performance at a group level cannot elucidate how the self-efficacy beliefs of any given individual relate to their *own* performance capacity. Calibration of self-efficacy is not concerned with whether individuals have high or low self-efficacy compared to others, but with whether an individual's self-efficacy beliefs over- or underestimate their own performance capacity. To elaborate, the self-efficacy beliefs of high-achieving student 'Alex' may be stronger than those of low-achieving student 'Andy', but they may simultaneously fall short of Alex's own capacity to perform, while in contrast, Andy may be over-efficacious. A correlation coefficient does not provide an accurate reflection of calibration in this case (Schraw, 2009a).

Miscalibration of academic self-efficacy may have negative impacts on academic outcomes. An over-efficacious student may be overly ambitious in choosing challenges and meet with failure as a result (Schunk & Pajares, 2004), or may alternatively underestimate the amount of effort or preparation that is required to successfully perform a task (Boekaerts

& Rozendaal, 2010; Hacker et al., 2000). Students may be less inclined to seek help and support (Zvacek, de Fátima Chouzal, & Restivo, 2015), and may exhibit complacency or carelessness (Vancouver, Thompson, Tischner, & Putka, 2002; Vancouver, Thompson, & Williams, 2001). Most commentary regarding miscalibration relates to concerns about how students with illusorily positive views lack the necessary realistic foundation from which to approach their learning in educational settings (e.g., Moore & Healy, 2008; Valentine, DuBois, & Cooper, 2004). However, under-efficaciousness may also be associated with inefficient academic self-regulation. An under-efficacious student may avoid attempting challenging tasks and thus prevent skill development, or may misallocate resources by over-studying (Boekaerts & Rozendaal, 2010; Schunk & Pajares, 2004). Under-efficacious students may also uncritically adopt unhelpful suggestions from others, or disengage because of a sense of self-doubt (Usher, 2016).

1.1 Calibration of students' academic self-efficacy beliefs

Our first research question was about calibration of students' academic self-efficacy beliefs. A small number of previous studies has considered the accuracy of mathematics self-efficacy beliefs with regard to performance outcomes in children and adolescents; in these studies, self-efficacy beliefs have generally been shown to exceed capacity to perform – that is, students are over-efficacious (e.g., Chen & Zimmerman, 2007; Chen, 2003; Pajares & Kranzler, 1995; Ramdass & Zimmerman, 2008). However, further research regarding calibration of academic self-efficacy is needed for a number of reasons. Firstly, the previous research has tended to focus on school-aged students, who are known to differ from adults in terms of factors influencing academic outcomes (Schneider & Preckel, 2017), including the strength of the self-efficacy/performance relationship (e.g., Richardson et al., 2012; Talsma et al., 2018). This suggests that more research with adult learners (e.g., university students) is necessary to determine whether similar patterns are evident in different educational settings.

Secondly, studies focusing on mathematics self-efficacy and performance involve very particular methods which are not applicable in many other natural learning tasks. In the studies cited above, a typical approach is to ask participants to look at mathematics problems briefly, and then rate their degree of confidence in correctly solving the problem. Immediately afterwards, participants solve the problems they were previously shown, or virtually identical problems. While this method provides a high degree of control and emulates the circumstances under which self-efficacy is known to be most predictive of learning outcomes (Bandura, 1997), there is some doubt as to whether such an approach is reflective of authentic learning experiences (Pieschl, 2009). When considering the university experience, written assignments and exams are common assessment methods, and these differ from previously studied judgements about mathematics capacity in a number of ways which may influence calibration findings. For example, participants have a genuine stake in the outcomes of these types of tasks, they apply a more complex range of self-regulated behaviours to perform them, and they are also influenced by different degrees of motivation (Gist & Mitchell, 1992; Hacker et al., 2008; Mengelkamp & Bannert, 2009). Another feature of the previous research regarding accuracy of mathematics self-efficacy is that studies have used *localised* judgements, which are made in relation to individual items (e.g., mathematics problems) and then averaged, whereas judgements referring to a whole task (i.e., *global* judgements, such as for an essay or exam) are more likely to occur naturally in educational settings (Pieschl, 2009). Finally, unlike tasks designed for the purposes of an experiment and administered in a single sitting – an approach which characterises much of the extant literature – assessment tasks undertaken as part of a course of study are also subject to different timing dynamics and greater uncertainty (Gist & Mitchell, 1992; Hacker et al., 2008; Mengelkamp & Bannert, 2009). Thus, research investigating the self-efficacy

judgements of student participants completing actual course assessments is likely to contribute to the ecological validity of research in this area.

An additional focus of the present study draws on previous literature which considered changing patterns of calibration depending on whether participants were stronger or weaker performers. Consistent with the "unskilled and unaware" effect (Kruger & Dunning, 1999), researchers report a tendency for stronger performers to underestimate their capacity, while weaker performers overestimate theirs (Bol, Hacker, O'Shea, & Allen, 2005; Hacker et al., 2008). Recent research suggests that a similar pattern of findings is evident when considering calibration of self-efficacy for school-aged music performers, with stronger music performers less confident in their abilities than weaker performers (Hewitt, 2015). However, it is not known whether this pattern would be evident in the self-efficacy calibration of university students studying traditional academic subjects. Thus, we also explored whether academically stronger students' self-efficacy beliefs underestimate their objective performance outcomes, and whether academically weaker students' self-efficacy beliefs exceed their objective performance outcomes.

1.2 Optimal level of academic self-efficacy

Our second key research question concerned the optimal level of self-efficacy in educational settings, which is a matter of discussion (e.g., Bandura & Locke, 2003; Vancouver & Kendall, 2006). Some researchers argue that self-efficacy beliefs which accurately reflect actual performance capacity will be of most benefit to performance (Stankov & Lee, 2017). Students who accurately judge what they can and cannot do are argued to be better able to adapt strategies effectively to the demands of a task (Boekaerts & Rozendaal, 2010). Chen (2003) reported that greater accuracy of school children's mathematics self-efficacy positively predicted subsequent mathematics performance.

However, researchers are far from unanimous in this regard. In a contrasting position, it is suggested that self-efficacy beliefs that exceed current capacity to perform are adaptive, motivating students to mobilise resources to increase performance above previous levels (Bandura, 1997). In this view, overestimation improves effort and persistence, and attempts to make students more realistic about their performance capacity are considered a dangerous enterprise (Pajares, 2006). In this vein, Pajares argues that a reach exceeding one's grasp should be encouraged – because one's sense of self-efficacy creates a self-fulfilling prophecy with respect to performance outcomes.

A third perspective is based on concerns that over-efficaciousness is instead potentially associated with a complacent attitude, whereby students may be content to 'coast' along in their studies without exerting effort or appropriately monitoring their performance (e.g., Dunlosky & Rawson, 2012; Zabrucky, 2010). Support for this is found in discrepancy-reduction theories of self-regulation (Carver & Scheier, 1982); researchers argue that overly strong self-efficacy beliefs may obscure any discrepancy between the current and desired state of learning, leading to reduced effort and poorer performance (Vancouver & Kendall, 2006). While collinearity precluded the use of self-efficacy bias in Chen's (2003) regression model predicting subsequent mathematics performance, she noted that self-efficacy bias and subsequent performance were negatively correlated – such that under-efficaciousness was associated with stronger subsequent performance.

Much of the commentary on what constitutes the optimal level of self-efficacy amounts to conjecture, because very few studies have directly addressed how calibration on one occasion relates to future behaviour (Bol & Hacker, 2012; Dunlosky & Rawson, 2012). The three positions outlined above would lead to three contrasting hypotheses: that better future performance could be predicted by either (1) accurate self-efficacy beliefs, (2) self-

efficacy exceeding current performance (over-efficaciousness), or (3) self-efficacy underestimating current performance (under-efficaciousness).

1.3 Summary

The present study aimed to explore university students' self-efficacy calibration for a range of academic outcomes forming part of their university studies (written assignments, exams, and overall subject performance), and to analyse how self-efficacy calibration relates to future academic performance.

Firstly, we explored self-efficacy accuracy and self-efficacy bias (see measures in the method section). Following previous findings in the literature, we hypothesised that:

1. Students' self-efficacy beliefs would be inaccurate overall (i.e., self-efficacy beliefs would deviate significantly in magnitude from objective performance outcomes).

2. Students' self-efficacy beliefs would be biased towards over-efficaciousness (i.e., self-efficacy beliefs would exceed objective performance outcomes when direction of the deviation is calculated).

3. Academically stronger students would be under-efficacious, while academically weaker students would be over-efficacious.

Secondly, by analysing two waves of data each for different types of academic tasks (written assignments and exams), we assessed how self-efficacy calibration (self-efficacy accuracy/bias) on one occasion related to subsequent performance on the same type of task. As discussed above, theoretical commentary suggests that multiple directional hypotheses are plausible (i.e., future academic performance outcomes may best be predicted by over-efficaciousness, under-efficaciousness, or accuracy of self-efficacy beliefs). For this reason, we did not stipulate a directional hypothesis with regard to the prediction of academic performance, but instead anticipated that:

4. Self-efficacy calibration (self-efficacy accuracy/bias) in relation to one academic task would predict performance outcomes on a similar subsequent task.

Exploration of the self-efficacy calibration of university students undertaking authentic performance assessments is needed to determine whether patterns of calibration are analogous to those identified in the previous literature. A greater understanding of the calibration of university students' self-efficacy beliefs with authentic academic performance outcomes including written assignments and exams, and subject-level performance, would be beneficial for educators and education policy makers, and may also contribute to theory development.

2 Method

2.1 Participants and procedure

Participants were 207 first-year undergraduate psychology students (152 female, mean age 25, age range 18–66 years) from an Australian university, who received course credit for participation. The data collection points are illustrated in Figure 1. Data were collected via online questionnaires and from institutional records over one academic semester plus the subsequent examination period in two cohorts (2014 [n=105] and 2015 [n=102]). After reading the study information and providing informed consent, participants completed a baseline questionnaire regarding self-efficacy for overall subject grade performance. Four identical questionnaires were subsequently completed within the five-day periods leading up to the submission of two written assignments and the completion of two multiple-choice exams. Ethics approval was obtained from the Human Research Ethics Committee of [withheld].

2.2 Measures

2.2.1 Academic performance. Academic performance in an introductory psychology course was measured following Australian standards, using grades with a possible range of 0 to 4 (0: fail [<50%], 1: pass [50-59%], 2: credit [60-69%], 3: distinction [70-79%], and 4: high distinction [≥80%]). Task-level academic performance was recorded as the grade achieved for the two written assignments and two exams completed over the course of the semester. The two written assignments were research reports written according to American Psychological Association guidelines. The two exams contained multiple-choice questions covering unit content. Together, these four tasks comprised 95% of the total assessment for the unit. Domain-level academic performance was recorded as the grade achieved for the subject overall, which comprised the task-level grades and one additional assessment task not considered in the present study. Thus, five sets of grades were recorded as academic performance measures for each participant: written assignment 1 and 2; exam 1 and 2; and subject grade. To ensure reliability, institutional policy provides that assessments are marked using rubrics and are moderated for consistency.

2.2.2 Self-efficacy. Based on recommendations in the literature, self-efficacy measures were designed to match performance outcomes as closely as possible (Bandura, 1986). As such, self-efficacy measures were tied to each of the five performance outcomes outlined above. Following Bandura's (2006) recommendations, participants were presented with each of the possible academic performance levels, in order of increasing difficulty, and were asked whether they were confident in their ability to achieve each one (this is a binary yes/no judgement; when "yes" responses are summed, this is referred to in the literature as self-efficacy magnitude; Lee & Bobko, 1994). Next, participants stated their level of confidence that they could attain each performance level (this is a 0–100% judgement; when percentages are added, this is referred to in the literature as self-efficacy strength; Lee & Bobko, 1994). A sample self-efficacy item demonstrating both of these elements is shown in

the Appendix. Following recommendations regarding the calculation of composite self-efficacy scores (Lee & Bobko, 1994), strength values (as decimals) were summed for each performance level that the participant indicated "yes" at magnitude level. E.g., a participant who responded "yes" they were confident in their ability to achieve a pass (with 100% confidence) and "yes" they were confident in their ability to achieve a credit (with 70% confidence), but "no" they were not confident in their ability to achieve a distinction (confidence 20%) would receive a score of 1.7 on the scale. Participants responding "no" to all binary decisions received a score of zero. Self-efficacy was measured on the same scale as performance, to facilitate the calculation of self-efficacy accuracy and self-efficacy bias scores as described below. By measuring self-efficacy and academic performance outcomes on the same scale, the calibration scores outlined below are comparable with those used in other calibration literature (see Schraw, 2009a; Schraw, 2009b).

      2.2.3 Self-efficacy calibration. Multiple measures of calibration exist, and these are interpreted differently; commentators recommend the use of multiple types of calibration indices so that findings can be clearly and comprehensively illustrated (Schraw, 2009b). In the present case, both self-efficacy accuracy and self-efficacy bias scores, outlined below, provided meaningful information regarding the calibration of university students' self-efficacy beliefs. Self-efficacy accuracy scores were required to test hypothesis 1, self-efficacy bias scores were required to test hypothesis 2, and both types of scores were required to test hypothesis 4 in the present study.

      2.2.3.1 Self-efficacy accuracy. Self-efficacy accuracy reflects the magnitude of the deviation between self-efficacy and performance, ignoring direction; it is a measure of "judgement precision" (see Schraw, 2009a, p. 35). All values are positive, with scores closer to zero reflecting greater accuracy. Five self-efficacy accuracy scores were calculated, with each reflecting the match between a single self-efficacy judgement and the corresponding

academic performance outcome. For each of the four task-based self-efficacy accuracy scores, the self-efficacy and performance data were taken from the same measurement occasion, as shown in Figure 1 (e.g., T2 written assignment1 self-efficacy was matched with T2 written assignment grades, resulting in self-efficacy accuracy for written assignment 1, and so on). For overall subject self-efficacy accuracy, T1 self-efficacy for the overall subject was matched with T6 overall subject grades.[1]

2.2.3.2 Self-efficacy bias. Self-efficacy bias is the signed deviation of the self-efficacy score from the academic achievement band score for each of the five performance outcomes; it is a measure of "the direction of judgement error" (see Schraw, 2009a, p. 35). A bias score of zero reflects perfect calibration, positive values indicate over-efficaciousness, and negative values indicate under-efficaciousness. Five self-efficacy bias scores were calculated, with each reflecting the signed match between one self-efficacy judgement and the corresponding academic performance outcome. Data points were as for self-efficacy accuracy.

2.3 Analyses

2.3.1 Calibration of students' academic self-efficacy beliefs

H1: A series of one-sample *t*-tests compared each of the five self-efficacy accuracy scores to a test value of zero (perfect calibration).

H2: A series of one-sample *t*-tests were conducted comparing each of the five self-efficacy bias scores to a test value of zero. However, it must be noted that as self-efficacy bias includes scores both above and below zero, a finding of no significant difference

---

[1] The self-efficacy accuracy measure outlined above is an example of absolute accuracy, as described in Schraw (2009a). Absolute accuracy reflects the match between an overall judgement of capacity and actual capacity. In contrast, relative accuracy measures tap into the match between capacity and performance on subtasks within an overall task (i.e., how well individuals are able to discriminate between what they know very well and what they know less well, Schraw, 2009b).

between mean bias and zero (perfect calibration) is not necessarily indicative of strong calibration. In this case, the presence of both over- and under-efficaciousness simultaneously could result in a mean score approaching zero. Thus, in order to supplement the bias t-tests with more easily interpretable results, proportions of participants who were over-efficacious, under-efficacious or calibrated were compared using Chi-square tests with a null hypothesis of equal distribution. In order to avoid deeming all students miscalibrated except those who were perfectly calibrated, a range around perfect calibration was established. As there is no widely accepted theoretical rationale for determining what scores constitute calibration, we took a pragmatic approach and followed Stankov and Lee (2014), such that participants whose bias scores were within 10% of calibration were considered calibrated.

H3: For each performance task, within-subjects t-tests were conducted to determine the difference between self-efficacy and performance at each of the five performance outcome levels, following previous calibration research (e.g., Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008). To supplement these analyses, we also prepared calibration plots. Calibration plots have the subjective judgement plotted on one axis and the objective outcome plotted on the other, in this case, comparing self-efficacy bias against a hypothetical line indicating perfect calibration. Unlike any single calibration co-efficient, calibration plots enable researchers to gauge the prevalence of both over- and under-efficaciousness, and they also provide an easily interpretable visual display of bias across performance levels (Hacker et al., 2008; Pieschl, 2009).

To reduce the risk of familywise error, the alpha level for the t-tests for hypotheses 1-3 was set at .01.

2.3.2 Optimal level of academic self-efficacy

H4: Two hierarchical regression analyses were conducted. A model of these analyses, using written assignment performance as an example, is shown in Figure 2. The

first analysis predicted performance on written assignment 2 (T3) using self-efficacy calibration measures (self-efficacy accuracy and self-efficacy bias) calculated from written assignment 1 (T2). The second analysis predicted performance on exam 2 (T5) using self-efficacy calibration measures (self-efficacy accuracy and self-efficacy bias) calculated from exam 1 (T4). Following Chen (2003), calibration measures from the first written assignment/exam tasks were used as predictors of performance from the second written assignment/exam tasks respectively, in order to avoid using non-independent data for the predictor and criterion variables. Task-based self-efficacy for the second written assignment/exam task was entered at step 1 of the hierarchical regression analyses, as measures of self-efficacy collected closer in time to the behaviour in question are known to be stronger predictors of that behaviour (Moriarty, 2014). That is, self-efficacy for written assignment 2 (T3) was used as a step 1 predictor for performance on written assignment 2 (T3). Similarly, self-efficacy for exam 2 (T5) was used as a step 1 predictor for performance on exam 2 (T5). Self-efficacy accuracy and self-efficacy bias, calculated from data for the first written assignment/exam tasks respectively, were entered at step 2. Cohort was also entered as a predictor variable at step 1 to account for variability associated with cohort differences. The three potential hypotheses regarding prediction of subsequent performance from self-efficacy calibration require measures that allow us to determine the effects of under-efficaciousness, over-efficaciousness, and accurate self-efficacy beliefs (see section 1.2). Consequently, both bias (over-/under-efficaciousness) and self-efficacy accuracy measures were included in regression models. Predictor variables were mean-centred.

3 Results

3.1 Preliminary analyses

Dropout analyses were conducted using logistic regression with attrition versus completion of all study questionnaires predicted from the study variables of baseline self-

efficacy and final grades. Participants completing all waves of data collection were more likely to have higher final grades (OR=1.04, *p*=.02).

A series of between-groups *t*-tests was conducted to compare the 2014/2015 cohort means on self-efficacy and performance variables to determine whether combining the two cohorts for analyses was appropriate. There was a significant difference between the two cohorts for grades on the second written assignment (2014: *M*=58.6, *SD*=14.1; 2015: *M*=64.2, *SD*=12.1). This was identified as an administrative error associated with marking procedures in 2015; therefore written assignment grades for the 2015 cohort were z-standardised and back-transformed to reflect the mean representing the institutional norm for that assessment. There were no other significant differences between the two cohorts.

3.2 Calibration of students' academic self-efficacy beliefs

Mean self-efficacy and academic performance for the five performance outcomes are shown in Table 2. Table 2 also shows the correlation between self-efficacy and performance for each assessment. Small-to-medium significant positive correlations were observed. Also shown in Table 2 are the mean self-efficacy accuracy and self-efficacy bias scores for the five performance outcomes. All self-efficacy accuracy scores were significantly different from zero (perfect calibration). With regard to self-efficacy bias, students were significantly over-efficacious regarding subject grades, and significantly under-efficacious regarding performance on the first exam. Mean bias for the other assessments did not differ significantly from zero because of roughly equal over-efficaciousness (positive scores) and under-efficaciousness (negative scores). This becomes clear when inspecting Figure 3, which shows the proportion of over-efficacious, calibrated, and under-efficacious participants for the five performance outcomes. Figure 3 shows that, for both written assignments and the final exam, over- and under-efficaciousness and calibration were roughly equally prevalent (χ^2 *p*s >.05), though there was a consistent slight tendency towards under-efficaciousness.

For the first exam, under-efficaciousness was pronounced, $\chi^2 = 121.30$, $p < .001$. Conversely, over-efficaciousness was prevalent in regard to overall subject performance, $\chi^2 = 136.46$, $p < .001$.

In the calibration plots (Figures 4–8) self-efficacy above the calibration line indicates over-efficaciousness, and below the line indicates under-efficaciousness. The plots indicate that, generally, weaker performers are over-efficacious while stronger performers are under-efficacious. The figures are annotated to show the results of the within-subjects t-tests comparing actual self-efficacy to performance at each grade level ($\alpha = .01$).

3.3 Optimal level of academic self-efficacy

Self-efficacy positively predicted performance, while self-efficacy bias negatively predicted performance (under-efficaciousness was associated with better performance), on both the written assignment and exam. Self-efficacy accuracy did not predict performance outcomes. Results of the hierarchical regression analyses are presented in Table 3. As shown, self-efficacy bias predicts subsequent performance incrementally to self-efficacy measured proximally to the task.

4 Discussion

This study examined the calibration of university students' self-efficacy beliefs with respect to actual performance outcomes. Overall, students' self-efficacy beliefs were poorly calibrated with actual capacity to perform. Self-efficacy beliefs were inaccurate overall for each of the performance outcomes. In terms of self-efficacy bias, under-efficaciousness was more prevalent at task level (written assignments and exams) and over-efficaciousness was more prevalent for overall subject outcome. Weaker performers tended to be over-efficacious, while the reverse was true for stronger performers. Self-efficacy bias on one occasion predicted subsequent performance on the same type of task, with under-efficacious

students showing better performance capacity. Accuracy of self-efficacy beliefs did not predict performance outcomes.

4.1 Calibration of students' academic self-efficacy beliefs

Analyses of self-efficacy accuracy showed that students' self-efficacy beliefs deviated significantly from their performance outcomes across all five measures. For individual academic tasks (written assignments and exams), under-efficaciousness was at least as prevalent as over-efficaciousness, accounting for more than a third of students for each task. This is in contrast with previous research that found the majority of students to be over-efficacious (Chen & Zimmerman, 2007; Chen, 2003; Ramdass & Zimmerman, 2008). As noted, the previous research focused on school-aged children, and it is possible that the age difference underlies the differences in findings reported here, especially given differences in self-efficacy levels and/or variations in the relationship between self-efficacy and performance across the lifespan (e.g., Davis-Kean et al., 2008). It is also plausible that the pressures of the authentic assessments for the participants in the present study contributed to these divergent findings, in contrast to previous studies which were conducted in laboratory settings. Further research is needed to determine whether these suggestions are supported. With regard to broader-level judgements that were more temporally distal, a different pattern emerged: more than two-thirds of participants' self-efficacy beliefs for overall subject grades exceeded their objective performance capacity, and less than ten percent were under-efficacious. This is more consistent with previous research on the accuracy of self-efficacy (Chen & Zimmerman, 2007; Chen, 2003) and research on calibration more generally (Hacker et al., 2000; Lin & Zabrucky, 1998)

Calibration plots and tests of differences between self-efficacy and performance at different performance levels illustrated a consistent tendency for weaker performers to be over-efficacious, and for stronger performers to be under-efficacious. While the magnitude

of this pattern varied across tasks, it was evident across all performance outcomes. While this pattern is consistent with previous research (Bol et al., 2005; Hacker et al., 2008), our findings suggest equal miscalibration for over- and underperformers whereas the previous studies suggest stronger miscalibration in underperformers.

4.1.1 Over-efficaciousness in weaker performers/under-efficaciousness in stronger performers.

A number of potential explanations for meta-cognitive overconfidence have been proffered, many of which are also plausible explanations for over-efficaciousness in weaker students. One example is the "unskilled and unaware" explanation, which suggests that poorer performers suffer a dual burden – the lack of capacity to perform well also manifests as a lack of insight into what is required to perform well, resulting in over-confidence (Kruger & Dunning, 1999). Similarly, weaker students' sense of efficacy may exceed their capacity to perform because they are unaware of where they are lacking – in this case, "unable and unaware". Other explanations based on motivated biases include the better-than-average effect (Krueger & Mueller, 2002), self-enhancement biases (Ehrlinger, Mitchum, & Dweck, 2016) and defensive self-deception (Stankov & Lee 2014).

Most previous research has discussed overconfidence, thus only little research attention has been paid to the potential underpinnings of under-confidence. However, several explanations are possible. For example, the self-efficacy beliefs of strong performers may be subject to defensive pessimism (Norem & Cantor, 1986) or "bracing for the worst", whereby people endeavour to protect themselves from the disappointment of negative results or feedback by lowering their outcome expectations (e.g., Taylor & Shepperd, 1998). This tendency is observed for self-relevant outcomes according to how detrimental an anticipated loss is perceived to be (Shepperd, Findley-Klein, Kwavnick, Walker, & Perez, 2000); thus, stronger students may be more likely to brace for loss given the greater likelihood that their self-concepts and self-esteem rest on strong academic performance (Higgins, 1987).

Recent qualitative research also suggests that individuals with stronger cognitive abilities may simply be more humble or cautious in their self-evaluations, preferring to provide lower judgements of confidence in their ability to perform (de Carvalho Filho, 2009). Students who underestimate their performance capacity report not wanting to "jinx" themselves and preferring to "play it safe" (Bol et al., 2005, p. 272). Strong students may also wish, for social reasons, to avoid appearing too competent or hard-working (Schunk & Pajares, 2004).

In observing over-efficaciousness in weaker performers and under-efficaciousness in stronger performers, in terms of task-based self-efficacy, it is also interesting to note that those who performed at an average level (i.e., those who achieved around a strong pass or credit grade) appeared to hold the most accurate self-efficacy beliefs. One might speculate that average students are less subject to the types of biases which influence the self-efficacy beliefs of those students at the more extreme ends of the academic performance continuum.

4.1.2 Over-efficaciousness for subject grades/under-efficaciousness for task-level assessments. While task-level self-efficacy calibration was not directly compared to subject-level in the present case, the statistics reported above converge on the conclusion that the grossest deviation from perfect calibration occurred at subject-level, and was in the direction of over-efficaciousness. This is consistent with other studies' findings of higher self-estimates at a broader level compared with a more specific level (Ackerman et al., 2002). The broader the domain under investigation, the more abstract self-judgements become, with the effect that they are more prone to influence from self-serving criteria (Dunning, Johnson, Ehrlinger, & Kruger, 2003).

A critical issue in terms of interpreting over-efficaciousness for subject grades and under-efficaciousness for task-level assessments relates to the timing of measurement. In the present study, self-efficacy for subject grades was measured close to the beginning of

semester; thus, subject-level self-efficacy judgements were made largely in the absence of relevant mastery information, a key source of self-efficacy beliefs (Bandura, 1997). A further timing issue is that the influences which may reduce overconfidence, such as defensive pessimism, are more likely to be at play as the potentially sobering outcome or feedback approaches, whereas temporally distant outcomes – such as overall grades – are more likely to be subject to unrealistic optimism (Sweeny & Krizan, 2013; Taylor & Shepperd, 1998). When students make self-efficacy judgements for broader-level outcomes that are distant in time, they may use inappropriate anchors for their self-efficacy beliefs and insufficiently adjust from that point (Bol & Hacker, 2012; Bol et al., 2005). For example, self-efficacy may be based on desired outcomes or aspirations (Serra & DeMarree, 2016) or norms such as average GPA of an institution (Clayson, 2005). In the case of task-level assessments, self-efficacy was measured close to the time of performance, potentially counteracting these influences.

While for three of the four task-based assessments, there was only a slight trend towards under-efficaciousness, in the case of the first exam, strong under-efficaciousness was evident. There are a number of plausible reasons for this. Calibration in this case may have been influenced by the "hard-easy effect" (Lichtenstein, Fischhoff, & Phillips, 1982). This effect is evident in research which suggests that individuals tend to underestimate the likelihood of success on tasks perceived to be easy, while overestimating the likelihood of success on tasks perceived to be hard (Merkle, 2009). Self-efficacy beliefs with regard to multiple-choice exams may be particularly subject to the hard-easy effect because it is known that students consider multiple-choice type assessments to be easy (e.g., Simkin & Kuechler, 2005). This effect may have been further exaggerated in this case; inspection of the mean performance ratings suggests that participants performed comparatively well on this task. It may be speculated that, by the time of the final exam, the experience of the first

exam as easy led to a partial reversal in the hard-easy effect, contributing to decreased under-efficaciousness for the second assessment of the same type. This suggestion, while plausible, awaits further empirical investigation.

4.2 Optimal level of academic self-efficacy

On all five measures, self-efficacy was significantly positively correlated with performance outcomes, with small-to-medium effects. This is consistent with previous findings in the literature. However, as discussed above, correlational analyses can only tell part of the story. While it is true that the strongest students' self-efficacy beliefs consistently exceeded those of the poorest students (reflected in the positive correlations and visible in the calibration plots), the self-efficacy beliefs of the strongest students also consistently fell short of their own capacity to perform, while weaker students by comparison were over-efficacious. The key issue is the point of comparison: low-achievers' self-efficacy beliefs are somewhat low when compared to high-achievers – but they are actually high when compared to their own performance outcomes. Thus, the positive correlation between self-efficacy and academic performance appears to obscure consistent trends in self-efficacy bias.

The regression results suggest that under-efficaciousness is the most adaptive calibration level, with self-efficacy bias tending towards under-efficaciousness being associated with stronger performance on subsequent analogous tasks. The effects noted were medium-to-large by convention. Self-efficacy accuracy did not significantly predict subsequent performance outcomes on either the written assignment or exam. This contrasts with findings that greater accuracy is related to better performance in mathematics for school-aged students (Chen, 2003). As we included both bias and accuracy in our analyses, we were able to pick up the distinction between over- and under-efficaciousness, both of which would have been reflected as non-directional inaccuracy in previous studies.

On face value, these findings are inconsistent with recommendations growing out of self-efficacy theory (Bandura, 1997; Pajares, 2006) that self-efficacy beliefs which exceed capacity are predictive of better outcomes. In contrast, they suggest that the motivation to reduce the discrepancy between the perceived current state of ability and the desired state of ability leads to increased effort and improved performance (e.g., Carver & Scheier, 1982; Vancouver & Kendall, 2006). This raises questions as to whether a reach exceeding one's grasp is an adaptive position, as some theorists contend (Pajares, 2006).

The above interpretation rests on an assumption regarding causality. The measurement of bias prior to performance provides some indication that it is not performance that directly influences bias (and we would not logically expect strong performance to directly result in under-efficaciousness). However, it seems perhaps most plausible that both self-efficacy bias and performance are related to a cocktail of other exogenous variables. For example, it may be that for strongly performing under-efficacious students, self-efficacy is not truly reflective of their belief in their ability to achieve, but is impacted by other motivations, such as a desire to "brace for loss", appear humble, or other motivations as discussed in section 4.1.1 above.

Calibration of self-efficacy presents itself as an important variable in the prediction of future performance, given that it explains variance in performance over and above that explained by self-efficacy measured immediately prior to the performance task, which is regarded as one of the strongest predictors of academic performance (Richardson et al., 2012; Robbins et al., 2004).

4.3 Limitations

The drop-out analysis identified that participants who completed all of the questionnaires were more likely to be better performers. This provides some evidence of systematic attrition in the present study, with poorer performers potentially under-

represented. In addition, as noted in the results section, there were differences in performance on the written assignment between the two cohorts included. Steps were taken to control for cohort differences.

With regard to over-efficaciousness in stronger performers and under-efficaciousness in poorer performers, we also note that there is a possibility that floor and/or ceiling effects may be present in our findings. Simply put, the most skilled performers have less room to be over-efficacious, while the poorest performers have less room to be under-efficacious (Krajc & Ortmann, 2008). However, recent modelling research suggests that these artefactual effects are unable to fully explain "unskilled and unaware" types of phenomena (Schlösser, Dunning, Johnson, & Kruger, 2013).

4.4 Directions for future research

A strength of the present study was the analysis of self-efficacy calibration across different task types and broader levels of outcome specificity not previously explored in self-efficacy calibration research. The findings highlight a potential lack of consistency in self-efficacy calibration depending on the performance outcome being investigated, and are suggestive of an interaction effect between ability level and performance task. Future research directly comparing calibration within-subjects across tasks and domains is warranted. Previous research suggests that interventions may improve the accuracy of mathematics self-efficacy beliefs in school-aged children (Ramdass & Zimmerman, 2008); further research into whether such interventions improve the accuracy of university students' self-efficacy beliefs is warranted. Furthermore, little is known about what differentiates over-efficacious and under-efficacious students (Ehrlinger et al., 2016). Identifying characteristics of biased students is potentially a fruitful avenue for further applied research.

In regard to the use of correlation coefficients and absolute accuracy as measures of calibration, it is evident from Table 2 and the calibration plots that very similar correlation

coefficients can show a very divergent distribution of over- and under-efficacious individuals across ability levels. Focusing only on accuracy would have obscured some important observations in the present case. We recommend that researchers exploring calibration carefully consider the range of measures available (Dunning & Helzer, 2014).

We speculated above that the differences between the present findings and previous research, in terms of the prevalence of under-efficaciousness reported here, may rest on the authentic nature of the assessments used for calibration and performance measures in the present study. An experimental approach comparing calibration for authentic assessment tasks and laboratory tasks in the same age group, and using the same calibration and performance measures, may provide evidence to support this supposition.

4.5 Conclusion

The findings of this study suggest that many university students' self-efficacy beliefs for academic performance are miscalibrated, with inaccuracy manifesting in both under- and over-efficaciousness. In a result that may surprise educators, educational researchers and policy makers, it was under-efficaciousness, rather than over-efficaciousness or accuracy of self-efficacy beliefs, which predicted future performance outcomes in this sample. We replicated the commonly reported finding that higher levels of self-efficacy were generally related to stronger performance outcomes; however, this finding cannot be interpreted in isolation. A key point is that, although low-achievers' self-efficacy beliefs did fall somewhat short of those of their high-achieving peers, they also were consistently higher than their own performance outcomes would warrant – that is, low-achievers were over-efficacious across all performance outcomes. Conversely, under-efficaciousness was prevalent in stronger performers. These findings cast some doubt on the notion of self-efficacy as self-fulfilling prophecy. In the simplest terms, we see that many students who believe they can, actually cannot, and many students who believe they cannot, actually can. Rather than

identifying under-efficaciousness as a risk factor for poor performance, our data support Shakespeare's assertion that modest doubt is a beacon of the wise.

The importance of increasing students' self-efficacy appears to be considered a *fait accompli* – this is evidenced by the many resources designed to help educators increase the self-efficacy of their students, which are available in both the academic domain (Ritchie, 2015; Sewell & St George, 2009) and in the popular press (Haskell, 2016; Kirk, n. d.). Yet, according to our results, it is unclear who would benefit from this approach, given that those under-efficacious students most in need of having their self-efficacy beliefs bolstered are those who are least in need of performance improvements. Meanwhile, the poorest performers appear to be in little danger of making self-debilitating judgements of their performance capacity. Further boosting the self-efficacy beliefs of low-achieving university students appears unlikely to be of benefit when these students appear to suffer no lack of self-efficacy when considering their actual performance capabilities. Instead, the over-efficaciousness observed in weaker performers puts them at risk of those potential negative impacts discussed earlier (Boekaerts & Rozendaal, 2010; Vancouver & Kendall, 2006), such as stopping studying before they are properly prepared for an assessment, or refraining from seeking academic support which is sorely needed.

A challenge for the future is to determine what constitutes an appropriate balance regarding approaches to self-efficacy in educational settings. On the one hand, theorists warn against translating findings such as those presented here into practical interventions designed to increase under-efficaciousness, citing the potential risk of disengagement (Vancouver & Kendall, 2006). Yet, we are unaware of any experimental studies confirming that this is what occurs. On the other hand, there is an increasingly observed tendency in educational settings to avoid providing negative feedback and constructive criticism (Dinham, 2010); a practice

which may be implicated in maintaining unrealistic self-efficacy beliefs which are not

helpful to students.

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*(4), 587-605.

    doi:10.1016/S0191-8869(01)00174-X

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*.

    Englewood Cliffs, NJ: Prentice-Hall, Inc.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan

    (Eds.), *Self-efficacy beliefs of adolescents* (Vol. 5, pp. 307-337). Greenwich, CT:

    Information Age.

Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited.

    Journal of Applied Psychology, 88(1), 87-99. doi:10.1037/0021-9010.88.1.87

Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration indices in order to

    capture the complex picture of what affects students' accuracy of feeling of

    confidence. *Learning and Instruction, 20*(5), 372-382.

    doi:10.1016/j.learninstruc.2009.03.002

Bol, L., & Hacker, D. J. (2012). Calibration research: Where do we go from here? *Frontiers

    in Psychology, 3*, 229: 221-226. doi:10.3389/fpsyg.2012.00229

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice,

    achievement level, and explanatory style on calibration accuracy and performance.

    The Journal of Experimental Education, 73(4), 269-290.

    doi:10.3200/JEXE.73.4.269-290

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for

    personality-social, clinical, and health psychology. *Psychological Bulletin, 92*(1),

    111-135. doi:10.1037/0033-2909.92.1.111

Chen, P., & Zimmerman, B. (2007). A cross-national comparison study on the accuracy of self-efficacy beliefs of middle-school mathematics students. *The Journal of Experimental Education, 75*(3), 221-244. doi:10.3200/JEXE.75.3.221-244

Chen, P. P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences, 14*(1), 77-90. doi:10.1016/j.lindif.2003.08.003

Clayson, D. E. (2005). Performance overconfidence: Metacognitive effects or misplaced student expectations? *Journal of Marketing Education, 27*(2), 122-129. doi:10.1177/0273475304273525

Davis-Kean, P. E., Huesmann, L. R., Jager, J., Collins, W. A., Bates, J. E., & Lansford, J. E. (2008). Changes in the relation of self-efficacy beliefs and behaviors across development. *Child Development, 79*(5), 1257-1269. doi:10.1111/j.1467-8624.2008.01187.x

de Carvalho Filho, M. K. (2009). Confidence judgments in real classroom settings: Monitoring performance in different types of tests. *International Journal of Psychology, 44*(2), 93-108. doi:10.1080/00207590701436744

Dinham, S. (2010). Self-esteem: Caution-do not over-inflate [online]. *Teacher: The National Education Magazine, August 2010*, 6-11.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271-280. doi:10.1016/j.learninstruc.2011.08.003

Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction, 24*, 58-61. doi:10.1016/j.learninstruc.2012.05.002

Dunning, D., & Helzer, E. G. (2014). Beyond the correlation coefficient in studies of self-assessment accuracy: Commentary on zell & krizan (2014). *Perspectives on Psychological Science, 9*(2), 126-130. doi:10.1177/1745691614521244

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*(3), 83-87. doi:10.1111/1467-8721.01235

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

Ehrlinger, J., Mitchum, A. L., & Dweck, C. S. (2016). Understanding overconfidence: Theories of intelligence, preferential attention, and distorted self-assessment. *Journal of Experimental Social Psychology, 63*, 94-100. doi:10.1016/j.jesp.2015.11.001

Gist, M. E., & Mitchell, T. R. (1992). Self-efficacy: A theoretical analysis of its determinants and malleability. *Academy of Management Review, 17*(2), 183-211. doi:10.5465/AMR.1992.4279530

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*(1), 160-170. doi:10.1037/0022-0663.92.1.160

Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in education: A focus on calibration. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 429-455). New York: Psychology Press.

Haskell, N. (2016). *Classroom strategies to improve student self-efficacy and learning outcomes.* Retrieved from http://www.pearsoned.com/education-blog/improve-self-efficacy-learning-outcomes/

Hewitt, M. P. (2015). Self-efficacy, self-evaluation, and music performance of secondary-level band students. *Journal of Research in Music Education, 63*(3), 298-313. doi:10.1177/0022429415595611

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94*(3), 319-340. doi:10.1037/0033-295X.94.3.319

Honicke, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review, 17*, 63-84. doi:10.1016/j.edurev.2015.11.002

Kirk, K. (n. d.). *Self-efficacy: Helping students believe in themselves.* Retrieved from https://serc.carleton.edu/NAGTWorkshops/affective/efficacy.html

Krajc, M., & Ortmann, A. (2008). Are the unskilled really that unaware? An alternative explanation. *Journal of Economic Psychology, 29*(5), 724-738. doi:10.1016/j.joep.2007.12.006

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*(2), 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*(6), 1121-1134. doi:10.1037/0022-3514.77.6.1121

Lee, C., & Bobko, P. (1994). Self-efficacy beliefs: Comparison of five measures. *Journal of Applied Psychology, 79*(3), 364. doi:10.1037/0021-9010.79.3.364

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art *Decision making and change in human affairs* (pp. 275-324): Springer.

Lin, L.-M., & Zabrucky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology, 23*(4), 345-391.

Mengelkamp, C., & Bannert, M. (2009). Judgements about knowledge: Searching for factors that influence their validity. *Electronic Journal of Research in Educational Psychology, 7*(17), 163-190.

Merkle, E. C. (2009). The disutility of the hard-easy effect in choice confidence. *Psychonomic Bulletin & Review, 16*(1), 204-213. doi:10.3758/pbr.16.1.204

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

Moores, T. T., Chang, J. C.-J., & Smith, D. K. (2006). Clarifying the role of self-efficacy and metacognition as predictors of performance: Construct development and test. *ACM SIGMIS Database, 37*(2-3), 125-132. doi:10.1145/1161345.1161360

Moriarty, B. (2014). Research design and the predictive power of measures of self-efficacy. *Issues in Educational Research, 24*(1), 55-66.

Norem, J. K., & Cantor, N. (1986). Anticipatory and post hoc cushioning strategies: Optimism and defensive pessimism in "risky" situations. *Cognitive Therapy and Research, 10*(3), 347-362. doi:10.1007/bf01173471

Pajares, F. (2002). Overview of social cognitive theory and of self-efficacy. Retrieved from https://www.uky.edu/~eushe2/Pajares/eff.html, 30 June 2017

Pajares, F. (2006). Self-efficacy during childhood and adolescence: Implications for teachers and parents. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 339-368). Greenwich, CT: Information Age.

Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology, 20*(4), 426-443. doi:10.1006/ceps.1995.1029

Pieschl, S. (2009). Metacognitive calibration—an extended conceptualization and potential applications. *Metacognition and Learning, 4*(1), 3-31. doi:10.1007/s11409-008-9030-4

Ramdass, D., & Zimmerman, B. J. (2008). Effects of self-correction strategy training on middle school students' self-efficacy, self-evaluation, and mathematics division learning. *Journal of Advanced Academics, 20*(1), 18-41. doi:10.4219/jaa-2008-869

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353-387. doi:10.1037/a0026838

Ritchie, L. (2015). *Fostering self-efficacy in higher education students*. London, UK: Palgrave Macmillan.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*(2), 261-288. doi:10.1037/0033-2909.130.2.261

Schlösser, T., Dunning, D., Johnson, K. L., & Kruger, J. (2013). How unaware are the unskilled? Empirical tests of the "signal extraction" counterexplanation for the dunning–kruger effect in self-evaluation of performance. *Journal of Economic Psychology, 39*, 85-100.

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin, 143*(6), 565. doi:10.1037/bul0000098

Schöber, C., Schütte, K., Köller, O., McElvany, N., & Gebauer, M. M. (2018). Reciprocal effects between self-efficacy and achievement in mathematics and reading. *Learning and Individual Differences, 63*, 1-11. doi:https://doi.org/10.1016/j.lindif.2018.01.008

Schraw, G. (2009a). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4*(1), 33-45. doi:10.1007/s11409-008-9031-3

Schraw, G. (2009b). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 415-429). New York: Routledge.

Schunk, D. H., & Pajares, F. (2004). Self-efficacy in education revisited. In D. M. McInerney & S. Van Etten (Eds.), *Big theories revisited* (Vol. 4, pp. 115-138). Greenwich, CT: Information Age Publishing.

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

Sewell, A., & St George, A. (2009). Developing efficacy beliefs in the classroom. *The Journal of Educational Enquiry, 1*(2).

Shepperd, J. A., Findley-Klein, C., Kwavnick, K. D., Walker, D., & Perez, S. (2000). Bracing for loss. *Journal of Personality and Social Psychology, 78*(4), 620-634.

Simkin, M. G., & Kuechler, W. L. (2005). Multiple-choice tests and student understanding: What is the connection? *Decision Sciences Journal of Innovative Education, 3*(1), 73-98.

Stankov, L., & Lee, J. (2014). Overconfidence across world regions. *Journal of Cross-Cultural Psychology, 45*(5), 821-837. doi:10.1177/0022022114527345

Stankov, L., & Lee, J. (2017). Self-beliefs: Strong correlates of mathematics achievement and intelligence. *Intelligence, 61*, 11-16. doi:10.1016/j.intell.2016.12.001

Stolp, S., & Zabrucky, K. M. (2009). Contributions of metacognitive and self-regulated learning theories to investigations of calibration of comprehension. *International Electronic Journal of Elementary Education, 2*(1), 7-31.

Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review, 12*(4), 437-475. doi:10.1023/A:1009084430926

Sweeny, K., & Krizan, Z. (2013). Sobering up: A quantitative review of temporal declines in expectations. *Psychological Bulletin, 139*(3), 702-724. doi:10.1037/a0029951

Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore i achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences, 61*, 136-150. doi:https://doi.org/10.1016/j.lindif.2017.11.015

Taylor, K. M., & Shepperd, J. A. (1998). Bracing for the worst: Severity, testing, and feedback timing as moderators of the optimistic bias. *Personality and Social Psychology Bulletin, 24*(9), 915-926. doi:doi:10.1177/0146167298249001

Usher, E. (2016). Personal capability beliefs. In L. Corno & E. M. Anderman (Eds.), *Handbook of educational psychology* (Vol. 3, pp. 146-159). New York: Routledge.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational psychologist, 39*(2), 111-133. doi:10.1207/s15326985ep3902_3

Vancouver, J. B., & Kendall, L. N. (2006). When self-efficacy negatively relates to motivation and performance in a learning context. *Journal of Applied Psychology, 91*(5), 1146-1153. doi:10.1037/0021-9010.91.5.1146

Vancouver, J. B., Thompson, C. M., Tischner, E. C., & Putka, D. J. (2002). Two studies examining the negative effect of self-efficacy on performance. *Journal of Applied Psychology, 87*(3), 506-516. doi:10.1037/0021-9010.87.3.506

Vancouver, J. B., Thompson, C. M., & Williams, A. A. (2001). The changing signs in the relationships among self-efficacy, personal goals, and performance. *Journal of Applied Psychology, 86*(4), 605. doi:10.1037/0021-9010.86.4.605

Zabrucky, K. M. (2010). Knowing what we know and do not know: Educational and real world implications. *Procedia: Social and Behavioral Sciences, 2*(2), 1266-1269. doi:10.1016/j.sbspro.2010.03.185

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, M. Zeidner, & P. Pintrich (Eds.), *Handbook of self-regulation* (pp. 13-39). California: Academic Press.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into practice, 41*(2), 64-70. doi:10.1207/s15430421tip4102_2

Zvacek, S. M., de Fátima Chouzal, M., & Restivo, M. T. (2015). Accuracy of self-assessment among graduate students in mechanical engineering. *Proceedings of the International Conference on Interactive Collaborative Learning (ICL), 2015*, 1130-1133. doi:10.1109/ICL.2015.7318192

Appendix

Self-efficacy was measured for each of the five academic performance outcomes in accordance with the example below. The model below was used for written assignments 1 (T2) and 2 (T3). When reporting self-efficacy for exam performance "upcoming written assignment" was replaced with "upcoming exam" for exam 1 (T4) and exam 2 (T5). Participants responded with regard to their "first-year psychology unit" as a measure of overall subject self-efficacy (T1). Calculation of a composite self-efficacy score based on responses to the questionnaire items below is detailed in the measures section of the method.

Step 1

Please indicate whether you believe you can achieve the assessment marks shown below for your upcoming written assignment.

	Yes	No
PP 50-59%	○	○
CR 60-69%	○	○
DN 70-79%	○	○
HD >80%	○	○

You will need to provide a response for every grade option, not just the highest grade you believe you are capable of achieving.

Step 2

Please rate how confident you are that you can achieve the assessment marks shown below. Rate your degree of confidence by recording a percentage from 0 to 100, where 0 means "cannot do at all", 50 means "moderately can do" and 100 means "highly certain can do".

PP 50-59%	
CR 60-69%	
DN 70-79%	
HD >80%	

You will need to provide a response for every grade option, not just the highest grade you believe you are capable of achieving. Please note that if you are e.g., 100% confident you can achieve a credit, please also note that you are 100% confident you can achieve a pass.

Table 1

Calculations of self-efficacy accuracy and self-efficacy bias for each of the five academic performance outcomes

	Self-efficacy accuracy
Written assignment 1	(Written assignment 1 self-efficacy [T2] *minus* written assignment 1 performance [T2])^2
Written assignment 2	(Written assignment 2 self-efficacy [T3] *minus* written assignment 2 performance [T3])^2
Exam 1	(Exam 1 self-efficacy [T4] *minus* exam 1 performance [T4])^2
Exam 2	(Exam 2 self-efficacy [T5] *minus* exam 2 performance [T5])^2
Subject grade	(Subject self-efficacy [T1] *minus* subject performance [T6])^2
	Self-efficacy bias
Written assignment 1	Written assignment 1 self-efficacy (T2) *minus* written assignment 1 performance (T2)
Written assignment 2	Written assignment 2 self-efficacy (T3) *minus* written assignment 2 performance (T3)
Exam 1	Exam 1 self-efficacy (T4) *minus* exam 1 performance (T4)
Exam 2	Exam 2 self-efficacy (T5) *minus* exam 2 performance (T5)
Subject grade	Subject self-efficacy (T1) *minus* subject performance (T6)

Table 2

Mean self-efficacy, grades, self-efficacy accuracy and self-efficacy bias, and correlations between self-efficacy and performance for each

performance outcome

	Written assignment 1	Written assignment 2	Exam 1	Exam 2	Subject grade
n	207	161	149	126	197
Self-efficacy	1.80 (.86)	1.84 (.85)	2.01 (.96)	1.88 (.90)	2.63 (.86)
Performance	1.82 (1.08)	1.76 (1.06)	2.96 (1.07)	1.96 (1.08)	1.62 (.92)
Correlation	.30***	.32***	.34***	.56***	.21**
Self-efficacy accuracy	0.93 (0.69)***	.91 (0.66)***	1.30 (0.81)***	0.80 (0.60)***	1.22 (0.93)***
Self-efficacy bias	-0.02 (1.16)	0.01 (1.13)	-1.04 (1.13)	-0.19 (0.99)***	1.05 (1.11)***

Note: Standard deviations are in parentheses; significance values for self-efficacy accuracy and self-efficacy bias reflect one-sample *t*-tests of the difference between mean

accuracy/bias and zero (perfect calibration) ** *p*<.01   *** *p*<.001

Table 3

Hierarchical regression results for prediction of written assignment and exam performance

	Step	Variables entered	β	R Square	R Square change
Written assignment (n=163)	1	Cohort	.046		
		Self-efficacy T3[a]	.288***	.084***	
	2	Cohort	.068		
		Self-efficacy T3[a]	.290***		
		SE Bias T2[b]	-.337***		
		SE Accuracy T2[b]	-.111	.212***	.128***
Exam (n=135)	1	Cohort	-.062		
		Self-efficacy T5[a]	.546***	.303***	
	2	Cohort	.091		
		Self-efficacy T5[a]	.690***		
		SE Bias T4[b]	-.402***		
		SE Accuracy T4[b]	.049	.453***	.150***

Note: SE = self-efficacy

[a] Self-efficacy was measured immediately prior to the performance outcome the second time the task was done (i.e., prior to written assignment 2 and exam 2). See Figure 2 for a visual representation, and the analyses section for rationale.

[b] Self-efficacy bias and self-efficacy accuracy were calculated based on measurements of self-efficacy and performance from the first time the task was done (e.g., self-efficacy for written assignment 1 and performance on written assignment 1). See Figure 2 for a visual representation the analyses section for rationale.

*** $p \leq .001$
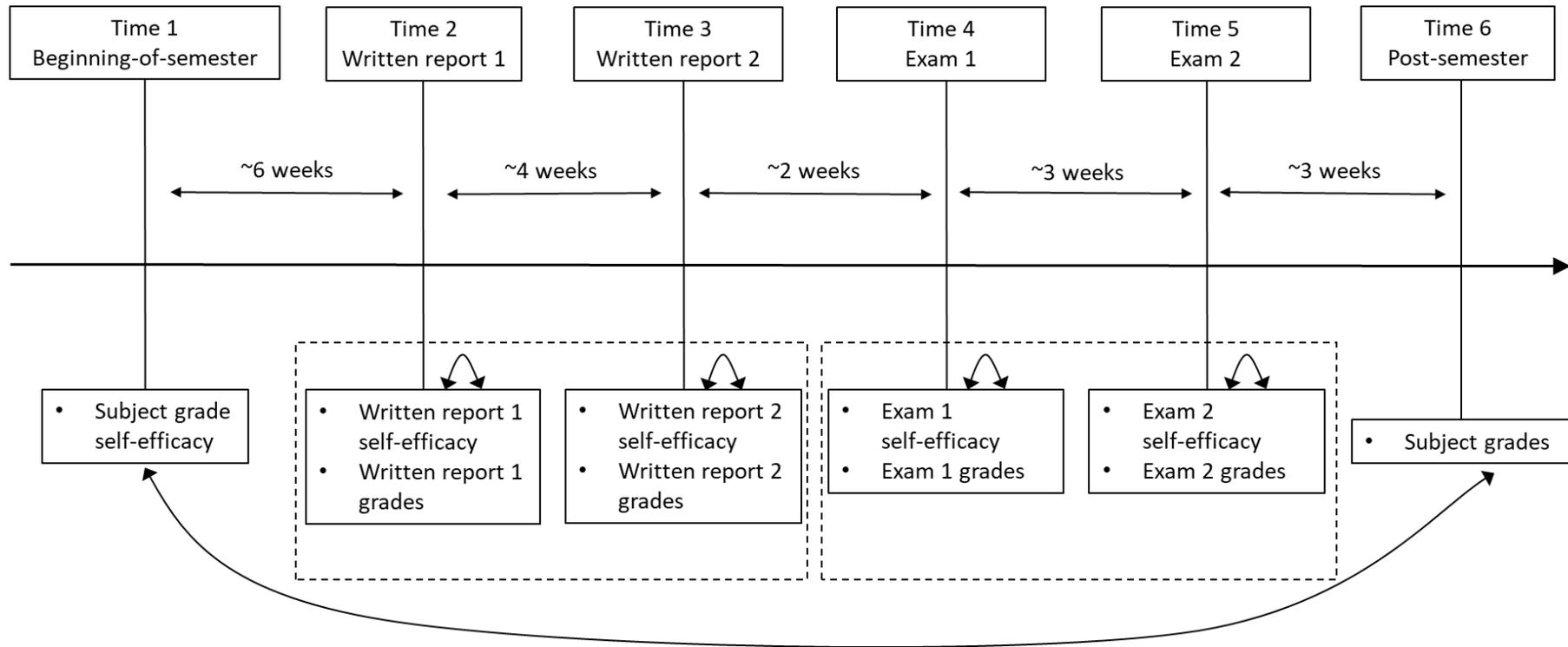
Figure 1. Study design. Regular boxes represent measurement occasion, academic task, and data collected. Curved arrows represent pairs of measures used for calculating self-efficacy accuracy and self-efficacy bias scores. Dashed boxes represent pairs of data waves used for regression analyses.
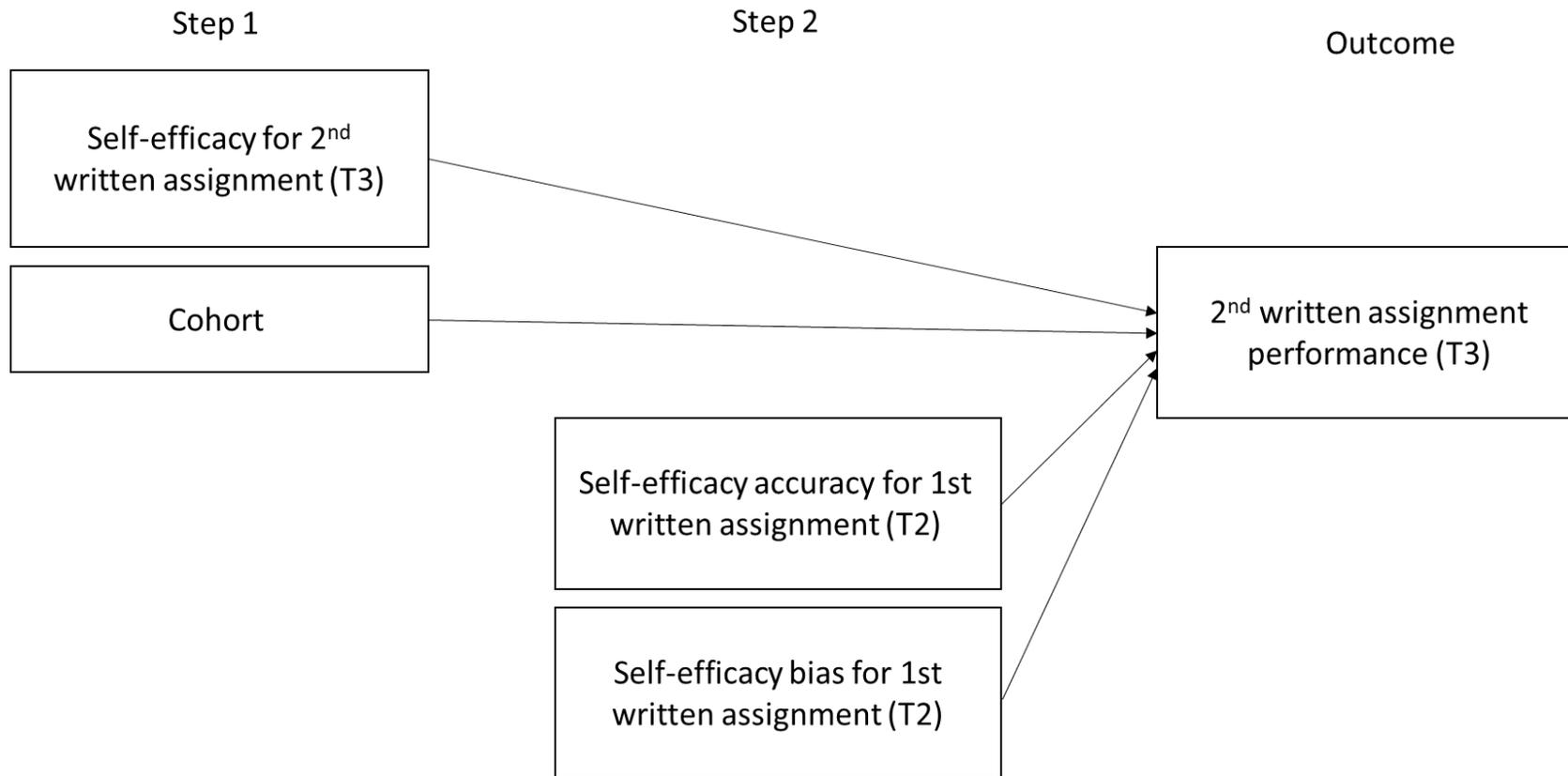
Figure 2. Model of regression analysis predicting performance on the second written assignment. The same model was used for performance on the second exam.
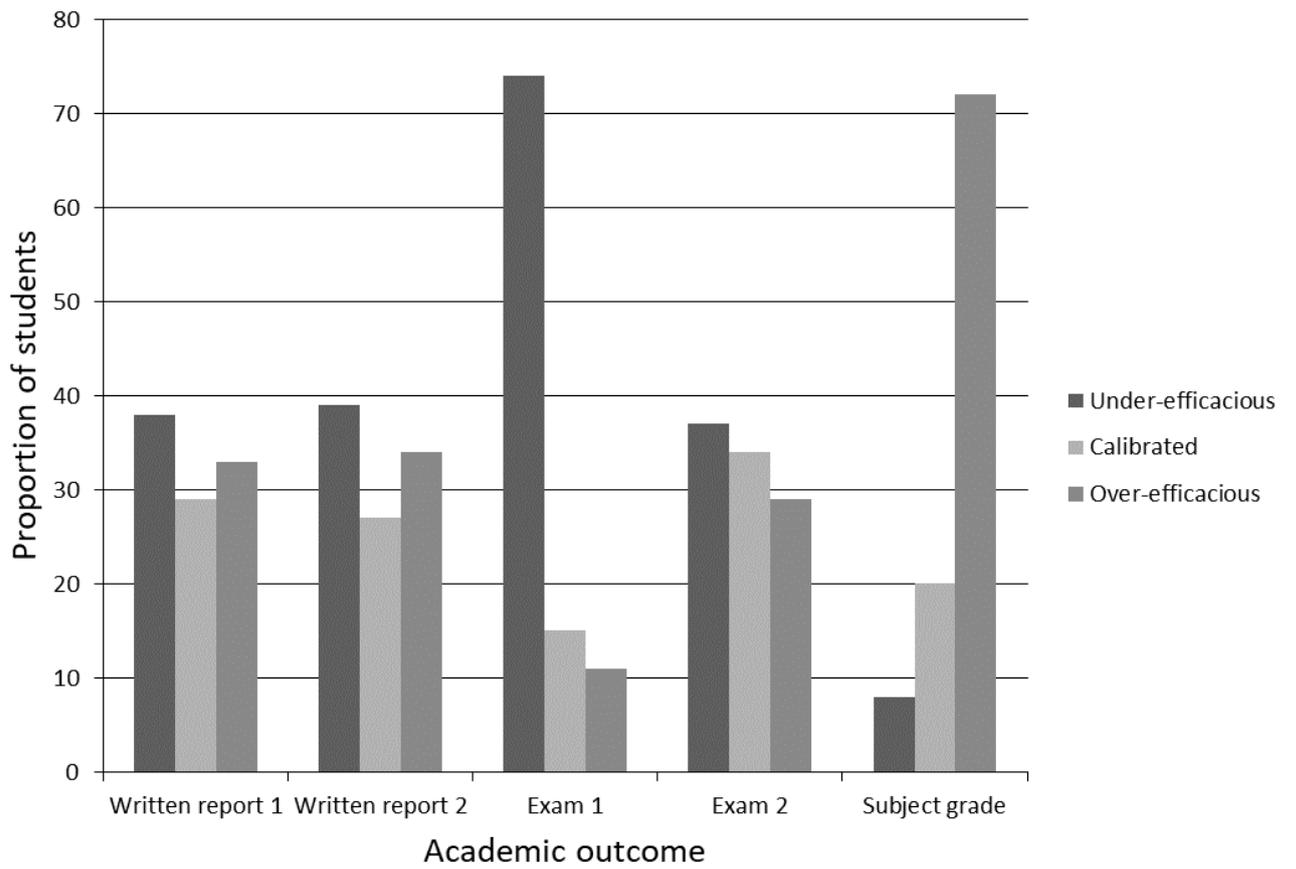
Figure 3. Self-efficacy calibration proportions by performance outcome
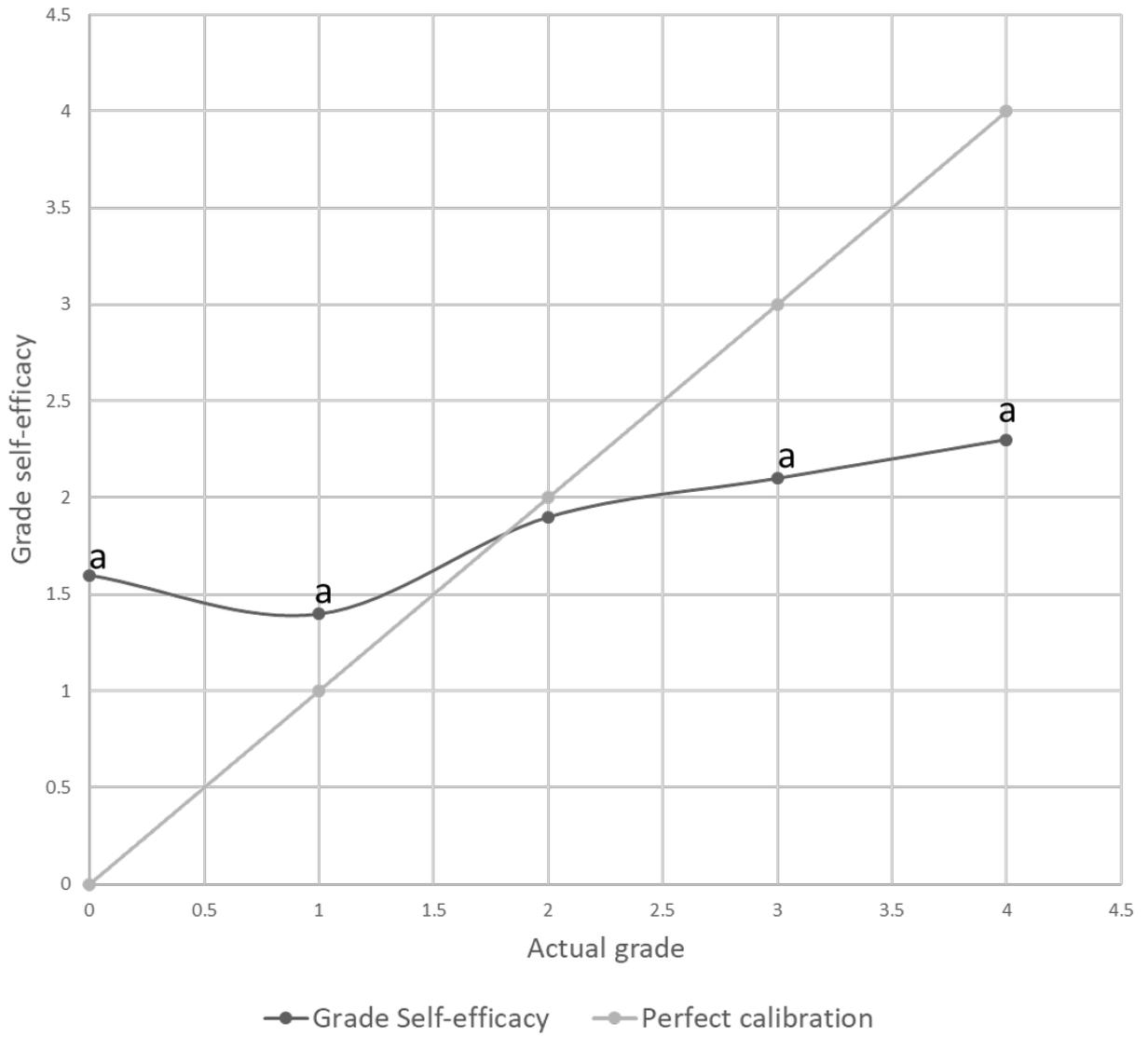
Figure 4. Self-efficacy calibration for written assignment 1

a = self-efficacy deviated from calibration at *p*<.001
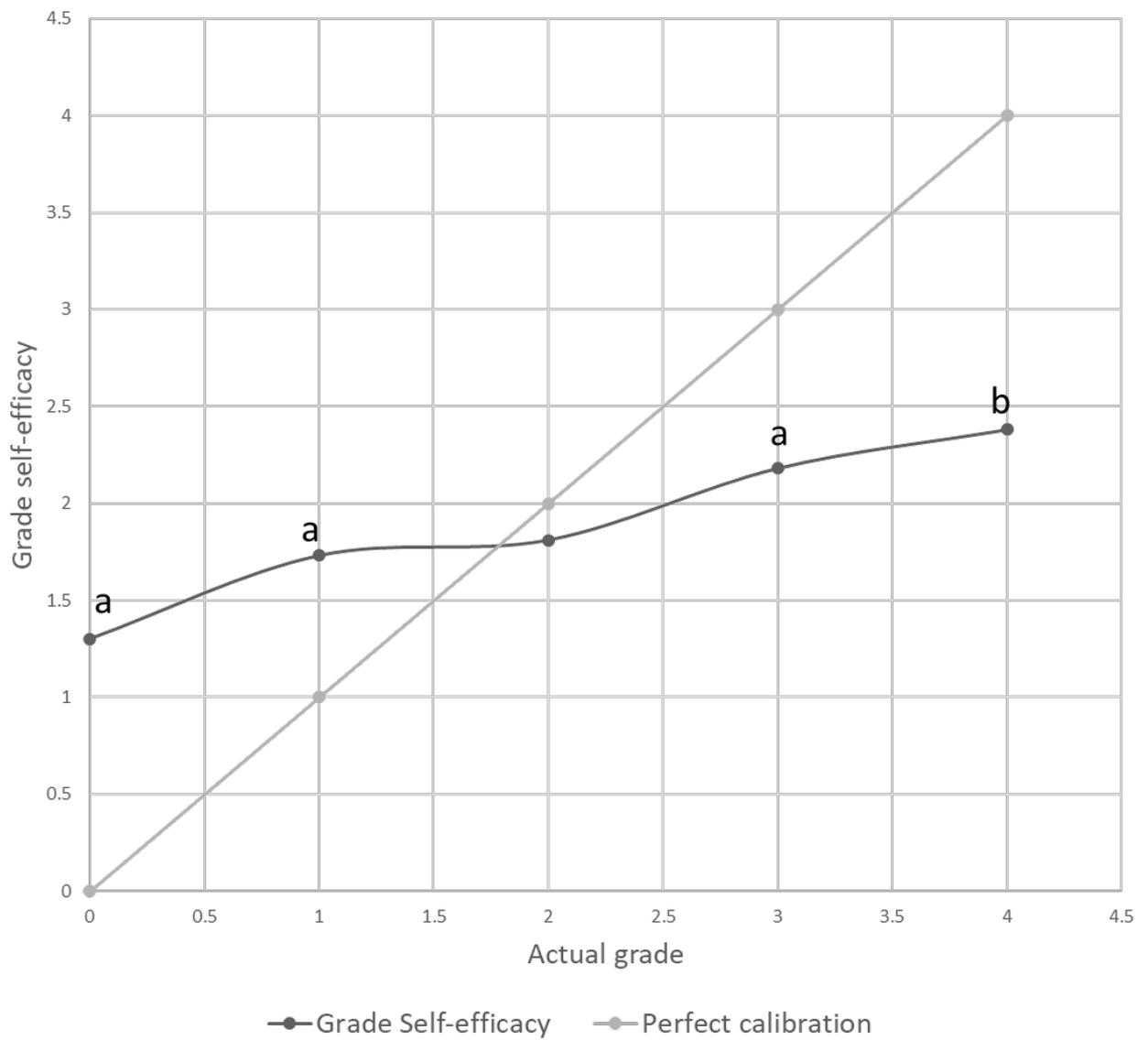
Figure 5. Self-efficacy calibration for written assignment 2

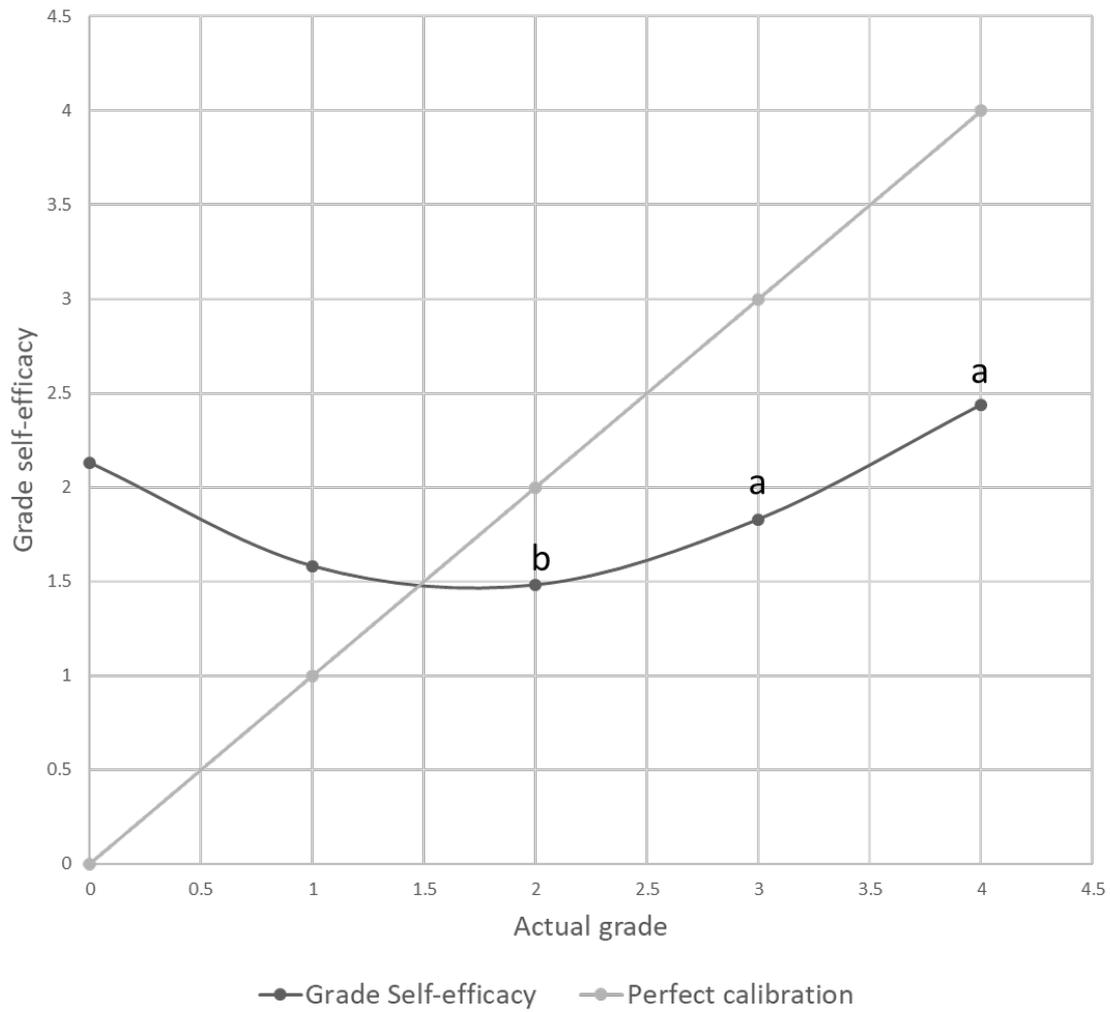a = self-efficacy deviated from calibration at *p*<.001, b = self-efficacy deviated from calibration at *p*<.01

Figure 6. Self-efficacy calibration for mid-semester exam

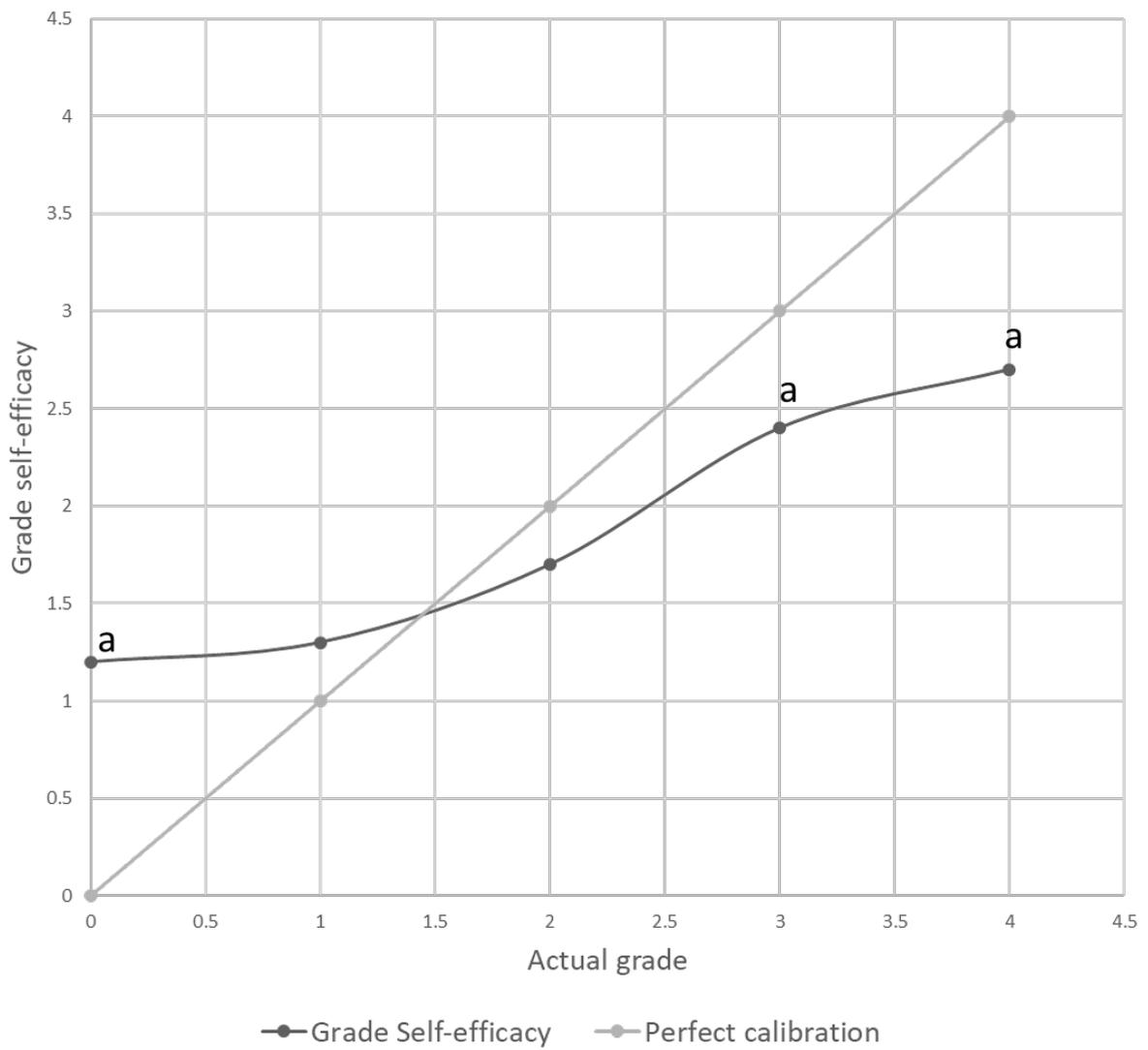a = self-efficacy deviated from calibration at $p<.001$, b = self-efficacy deviated from calibration at $p<.01$

Figure 7. Self-efficacy calibration for end-of-semester exam

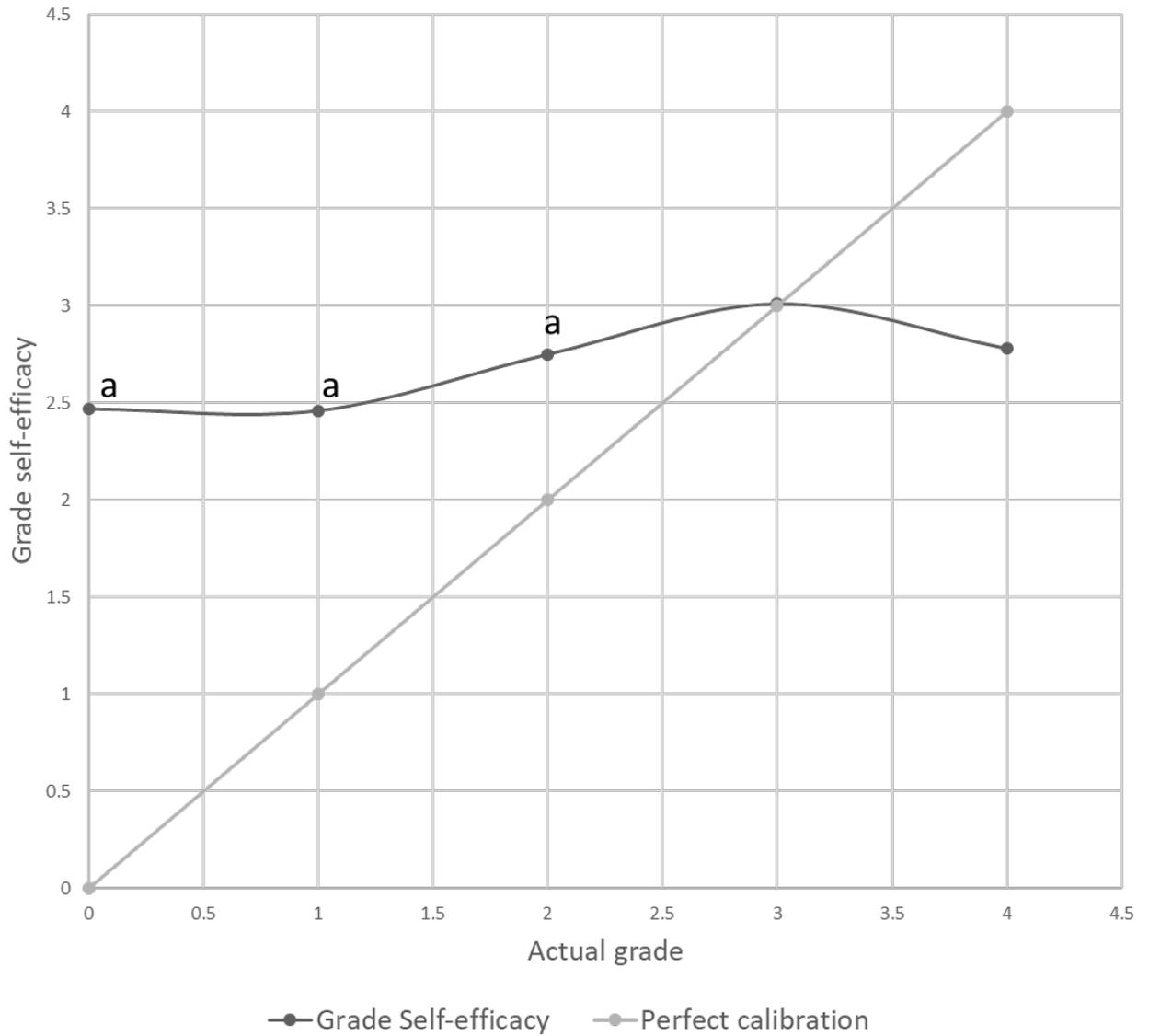a = self-efficacy deviated from calibration at *p*<.001

Figure 8. Self-efficacy calibration for overall subject grade

a = self-efficacy deviated from calibration at *p*<.001