First-Year Students' Academic Self-Efficacy Calibration: Differences by Task Type, Domain Specificity, Student Achievement Level, and Over Time

Kate Talsma
University of Tasmania, Australia
Kimberley Norris
University of Tasmania, Australia
Benjamin Schüz
University of Bremen, Germany

Abstract

This research explored whether academic self-efficacy calibration (the match between self-efficacy beliefs and academic outcomes) in first-year psychology students (n=197) differed as a function of task type (written assignment/multiple-choice exam), domain specificity (task level/subject level), over time (mid-semester/end of semester) and according to student achievement level (high achievers/low achievers). Lower-achieving students were overconfident across both the written assignment and the exam, while higher-achieving students were accurately calibrated on both tasks. The subject-level calibration of lower-achieving students improved between mid-semester and the end of semester (though students remained overconfident). Higher-achieving students' subject-level calibration remained stable over the semester, and they were about half as overconfident as the lower-achieving students. Both groups of students were more overconfident at subject-level than at task-level overall. On the whole, overconfidence was prevalent, especially for low achievers, and at subject level. Findings suggest that a one-size-fits-all approach to self-efficacy is unlikely to be beneficial for all learners.

Keywords: Student experience; academic performance; self-efficacy; first-year students.

Introduction

Self-efficacy is a "can-do" belief about one's capacity to achieve future outcomes (Bandura, 1997). Academic self-efficacy is considered one of the primary non-intellectual determinants of academic performance (Schunk & Pajares, 2004). Meta-analyses consistently show that self-efficacy is positively related to academic achievement (e.g., Honicke & Broadbent, 2016; Richardson et al., 2012). As a result, the position that high self-efficacy is beneficial, and low self-efficacy is not, is so widespread as to be virtually passé (e.g., Artino, 2012). However, some researchers argue that the situation is more complex than this, and warrants further investigation (Talsma et al., 2018; Vancouver, 2012).

There is growing interest in the accuracy, or *calibration*, of students' beliefs about their academic abilities (Hattie, 2013). Calibration refers to the match (or discrepancy) between subjective judgements and the corresponding objective reality. Self-efficacy is an important construct for investigation within a calibration paradigm, as it is a key component of prominent self-regulation theories (e.g., Zimmerman & Moylan, 2009). It has been argued that effective monitoring and self-regulation are not

possible without accurate knowledge of one's present and target states (Bol & Hacker, 2012; Hattie, 2013). Research indicates that learners' self-efficacy beliefs are often miscalibrated (Talsma, Schüz, et al., 2019; Zimmerman et al., 2011). Self-efficacy beliefs might exceed objective capacity (i.e., the student is overconfident, or has a *positive self-efficacy bias*) or fall short of it (i.e., the student is underconfident, or has a *negative self-efficacy bias*). Students may be accurately calibrated, in that there is a good match between their self-efficacy beliefs and their academic performance outcomes.

A deeper understanding of calibration is critical to effective learning and teaching, because students' calibration levels affect ongoing cognitive control processes and self-regulation (e.g., Dunlosky & Thiede, 2013). Importantly, self-efficacy calibration can predict additional variance in future academic outcomes beyond that predicted by self-efficacy alone (Talsma, Schüz, et al., 2019). Overconfident students may cease studying prematurely, refrain from seeking help or trying new strategies, and ultimately, perform more poorly (e.g., Boekaerts & Rozendaal, 2010). While underconfidence is less common (e.g., Zabrucky, 2010), it has been observed, particularly in stronger students (e.g., Kruger & Dunning, 1999; Zimmerman et al., 2011). Underconfidence does not tend to be detrimental to performance outcomes (Talsma, Schüz, et al., 2019), but underconfident students may over-invest resources studying materials already mastered (Bol & Hacker, 2012; Hattie, 2013).

The purpose of this paper is to explore how academic self-efficacy calibration varies across a range of dimensions specific to the learner and the learning context. Knowing the circumstances under which students are better or worse at calibrating their self-efficacy beliefs may assist educators to understand factors impacting on students' motivation and achievement. This knowledge may also inform research regarding the potential sources of bias in self-efficacy beliefs, and whether interventions to enhance calibration are beneficial, and for whom. The following factors are proposed to influence self-efficacy calibration.

Task Type

Task-specific cues are likely to influence the accuracy of students' beliefs about their capacity to perform (Bol et al., 2005; Gist & Mitchell, 1992). Students may demonstrate different degrees of accuracy in their judgements about different types of tasks, such as those involving different text genres (Lin & Zabrucky, 1998) or different types of maths problems (Boekaerts & Rozendaal, 2010). In the present study, we compare calibration of self-efficacy for a written assignment with calibration of self-efficacy for a multiple-choice exam. The likely direction of any difference between these tasks is unclear. Some research suggests that, consistent with the "hard-easy" effect, people are less accurate (more overconfident) with regard to complex tasks than simple tasks (Boekaerts & Rozendaal, 2010; Lichtenstein et al., 1982). Thus, we might expect students to be more overconfident on a written assignment than a multiple-choice exam. Conversely, some researchers report that participants tend to better judge their abilities with regard to more complex open-ended questions, than simpler multiple-choice items (Bol & Hacker, 2001; de Carvalho Filho, 2009), owing to greater activation of content knowledge and depth of processing required for the more complex task. Thus, we expected that self-efficacy calibration would differ across task types, but did not specify a direction.

Domain Specificity

The self-efficacy literature distinguishes between task- and domain-level specificity (e.g., Pajares & Usher, 2008). In learning settings, task-level self-efficacy may refer to a single academic activity such as an essay or a test, while domain-level self-efficacy refers more broadly to a subject area or course. Scores on more specific self-efficacy measures are generally more strongly correlated with task performance than scores on less specific measures (Bong, 2001). In meta-analyses, measurement specificity moderates the relationship between self-efficacy and performance, with more specific measures showing stronger associations (e.g., Talsma et al., 2018). It is possible that the weaker relationship between self-efficacy and performance at domain level is underscored by consistent self-efficacy biases at that level (Ackerman et al., 2002). Self-judgements become more abstract as the domain in question becomes broader; they are more prone to influence from self-serving criteria (Dunning et al., 2003). Thus, it was expected that self-efficacy bias would be higher (students would be more overconfident) at subject level than at task level.

Time

As mastery experience is a key source of self-efficacy beliefs (Bandura, 1997; Usher & Pajares, 2008), miscalibration may decrease over time as experience is gained. Support for this position comes from previous research showing that self-efficacy

increases in its predictive utility as courses progress (Gore Jr, 2006). This trend is perhaps even more likely in the case of first-year undergraduates, whose early self-efficacy judgements may be anchored to inappropriate sources of information in the absence of domain-relevant mastery experience (e.g., Bol et al., 2005). Also, as the end of semester approaches, the common trend towards unrealistic optimism may be subdued by the growing proximity of the academic outcome in question (Taylor & Shepperd, 1998). Thus, it was expected that calibration of self-efficacy and overall subject performance would be more accurate at the end of semester than mid-semester.

Ability

Lower-performing students tend to demonstrate the greatest degree of miscalibration, typically in the direction of overconfidence, while higher-performing students are more accurate, and may express a degree of underconfidence (e.g., Bol et al., 2005; Kruger & Dunning, 1999; Talsma, Schüz, et al., 2019). It is argued that both cognitive and motivational biases underlie these differences. The "unskilled and unaware" argument suggests that those with lower performance capacity in a domain of functioning simultaneously lack awareness around the capacity required to perform, such that they are unable to accurately identify their own level of (in)competence (e.g., Fritzsche et al., 2018; Kruger & Dunning, 1999). At the same time, stronger students may be driven to under-report their capacities for several reasons – whether because of defensive pessimism or "bracing for the worst" (Norem & Cantor, 1986; Taylor & Shepperd, 1998), because of a desire to be cautious or "play it safe" (Bol et al., 2005; de Carvalho Filho, 2009), or in order to present one's capabilities according to perceived social requirements (Schunk & Pajares, 2004). We included this ability factor across all analyses and anticipated seeing a consistent effect of ability across contexts such that lower-performing students would be more overconfident than higher-performing students.

In sum, it was hypothesised that:

H1: Mean self-efficacy bias would differ across two assessment tasks (written assignment and multiple-choice exam), with lower-performing students showing higher self-efficacy bias scores (more overconfidence) than higher-performing students across both tasks.

H2: Mean self-efficacy bias at subject level would be lower at the end of semester (less overconfidence) than at mid-semester, with lower-performing students showing higher self-efficacy bias (more overconfidence) than higher-performing students on both occasions.

H3: Students would be more accurately calibrated at task level than subject level overall, with lower-performing students showing higher self-efficacy bias (more overconfidence) than higher-performing students at both levels of specificity.

Method

Participants and Procedure

Participants were 197 first-year undergraduate students (144 female, M_{age}=25 years, range 18 – 67 years) studying at a regional Australian university. Participants were all enrolled in the same introductory psychology class, and received course credit for their participation. Ethics approval was obtained from the Human Research Ethics Committee of the University of Tasmania. Self-efficacy data were collected via an online survey on two occasions during the semester, within the five-day periods prior to the submission of a written assignment (week 6) and a multiple-choice exam (week 14). On the first occasion, students provided their self-efficacy beliefs for the upcoming written assignment and the subject overall. On the second occasion, students provided self-efficacy judgements about the upcoming exam, and again about overall subject outcomes. Of the 197 participants at time 1, 126 completed the time 2 measures. With permission from participants, three grades were collected from institutional records (written assignment, multiple-choice exam, and subject overall).

Measures

Self-efficacy bias (SE bias, i.e., calibration) was calculated as the signed deviation between the students' self-efficacy (SE) scores and their academic performance (AP) scores: AP was subtracted from SE. AP and SE measures are described below. Overconfidence was reflected in positive SE bias scores, in the case of participants whose SE scores exceeded their AP scores. Underconfidence was reflected in negative self-efficacy bias scores, when participants' AP scores exceeded their SE scores. A

bias score of zero reflected accurate calibration. Two task-level SE bias scores were calculated (written assignment grade was subtracted from written assignment self-efficacy, and multiple-choice exam grade was subtracted from exam grade self-efficacy) and two subject-level SE scores were calculated (overall subject grade was subtracted from both time 1 and time 2 subject self-efficacy).

Academic performance (AP) was operationalised as grade data collected from institutional records at the end of semester. AP was measured following Australian standards as the grade achieved for a written assignment, multiple-choice exam, and the subject overall, 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%], respectively). In the analyses outlined below, the IV of high/low ability was determined by a median split of AP scores for the subject overall.

Self-efficacy (SE) was measured by self-report as confidence in ability to achieve grades measured on the same scale as academic performance, again presented in order of increasing difficulty.

Analyses

A series of 2 x 2 mixed factorial ANOVAs were conducted to test the study hypotheses:

H1: 2 (task: assignment, exam) x 2 (ability: high/low)

H2: 2 (subject: T1, T2) x 2 (ability: high/low)

H3: 2 (specificity: task, subject) x 2 (ability: high/low).

We also used the 95% confidence intervals of the means from the above analyses to determine calibration with reference to the zero point (accurate calibration). Overconfidence was evident if mean self-efficacy bias scores were positive and 95% confidence intervals did not cross zero. Underconfidence was reflected in mean self-efficacy bias scores that were negative with 95% confidence intervals not crossing zero. If 95% confidence intervals included positive and negative valence and included the zero point representing calibration, this indicated accurate calibration).

Results

Descriptive statistics are provided in Table 1, along with correlations between self-efficacy and academic performance, which were medium-to-large across measurements. Proportions of participants at different calibration levels across tasks are shown in Table 2.

Table 1

Means and standard deviations (in parentheses) for main study variables and correlations between self-efficacy and grades for each academic outcome

	Time 1 (*n*=197)		Time 2 (*n*=126)	
	Written report	Subject	Exam	Subject
Self-efficacy	2.411 (0.925)	2.711 (0.841)	2.468 (0.985)	2.468 (0.935)
Grade	1.888 (1.043)	n/a	2.064 (1.151)	1.619 (0.922)
Calibration	0.523 (1.150)	1.091 (1.055)	0.405 (1.052)	0.730 (0.862)
Correlation	.330**	.319**	.540**	.610**

** p<.01
Note: There is only one subject grade. Time 1 subject calibration and correlation values were calculated using end of semester subject grade and mid-semester subject self-efficacy. Time 2 subject calibration and correlation values were calculated using end of semester subject grade and end of semester subject self-efficacy.

Table 2

Proportions of overconfident, underconfident and accurately calibrated participants

		Time 1 (*n*=197)		Time 2 (*n*=126)	
		Written report	Subject	Exam	Subject
Higher achievers	Overconfident	37.7%	58.2%	29.6%	52.2%
	Calibrated	34.7%	35.7%	42.3%	42.3%
	Underconfident	27.6%	6.1%	28.2%	5.6%
Lower achievers	Overconfident	61.7%	88.9%	70.9%	74.6%
	Calibrated	29.3%	8.1%	14.5%	20%
	Underconfident	9.1%	3%	14.5%	5.5%

Hypothesis 1 was partially supported. There was no main effect of task: self-efficacy bias for the written assignment (*M*=0.499, 95%CI[0.322, 0.675]) did not differ significantly from self-efficacy bias for the multiple-choice exam (*M*=0.460, 95%CI[0.288, 0.631]), $F(1,124)=.130$, $p=.719$. However, there was a main effect of ability as anticipated: the self-efficacy bias scores of lower-performing students (*M*=0.845, 95%CI[0.640, 1.051]) were higher than those of higher-performing students (*M*=0.113, 95%CI[-0.068, 0.294]), $F(1,124)= 28.062$, $p<.001$, $d=0.952$, 95%CI[0.581, 1.323]. Task-specific means and 95% confidence

intervals for each group are shown in Figure 1. The 95% confidence intervals indicate that, across tasks, lower-performing students were overconfident, while higher-performing students did not differ significantly from accurately calibrated on average. No interaction effect was hypothesised or found; $F(1,124)=1.438$, $p=.233$.

Figure 1

Self-efficacy Bias Across Task Type and Achievement Level (n=126)
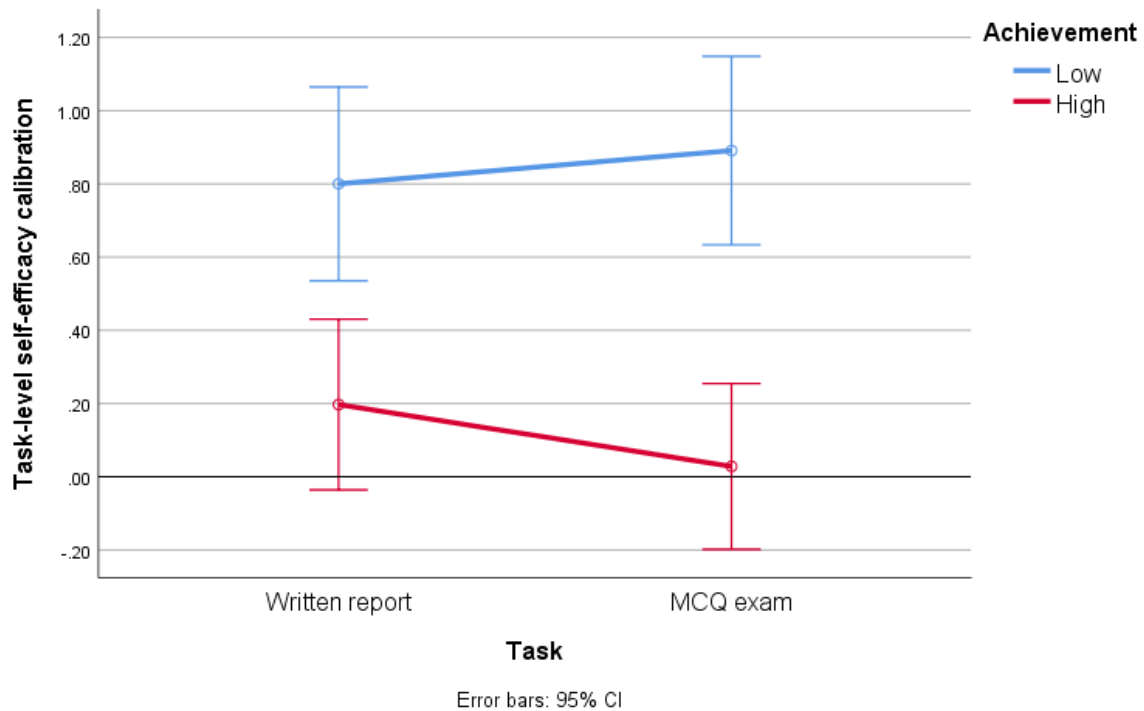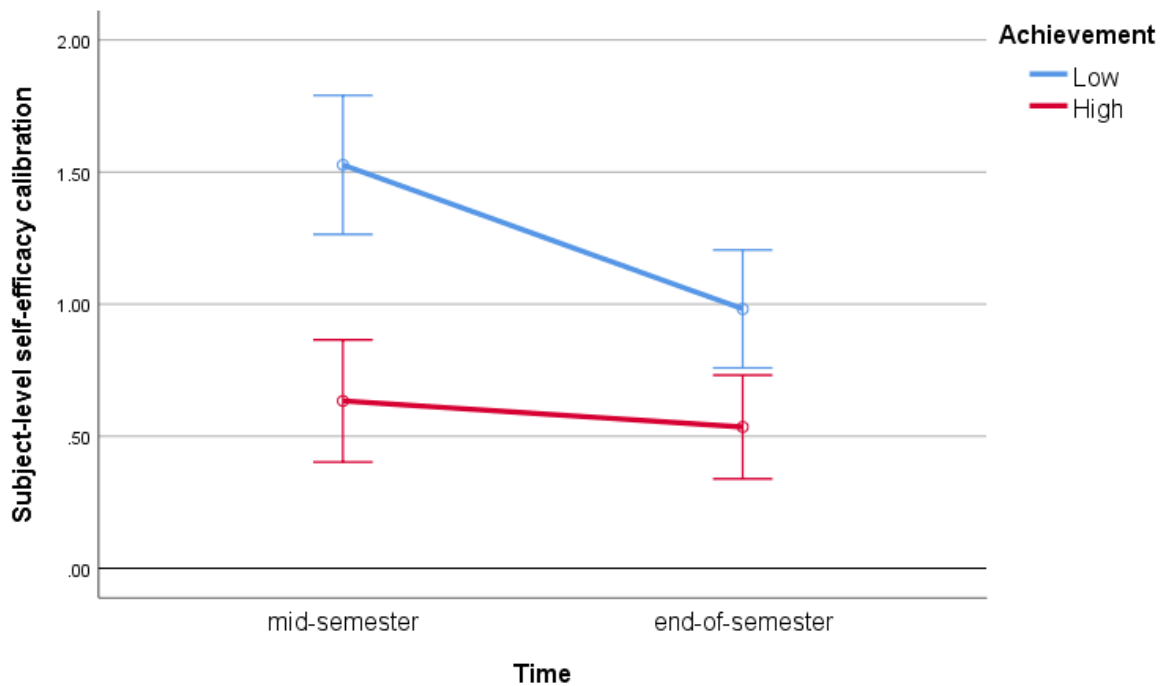

Note: Error bars are 95% confidence intervals. Values on the Y axis indicate the average difference between self-efficacy and academic performance as proportions of a grade band (zero means accurate calibration, positive values mean overconfidence, and negative values mean underconfidence).

Consistent with hypothesis 2, there was a main effect of measurement occasion for subject-level self-efficacy bias, which was significantly higher (showing greater overconfidence) at mid-semester ($M=1.081$, 95%CI[0.906, 1.256]) than end of semester ($M=0.759$, 95%CI[0.610, 0.907]), $F(1,124)=25.733$, $p<.001$, $d=0.390$, 95%CI[0.140, 0.639]. Again, there was a main effect of ability, such that the subject-level self-efficacy bias scores of lower-performing students ($M=1.255$, 95%CI[1.030, 1.479]) exceeded those of higher-performing students ($M=.585$, 95%CI[0.387, 0.782]), $F(1,124)=19.622$, $p<.001$, $d=0.794$, 95%CI[0.428, 1.159]. The 95% confidence intervals for these means indicate that, at subject-level, lower-performing students were overconfident by at least one grade band, while higher-performing students were overconfident by approximately half a grade band on average. Subject-level means and 95% confidence intervals for each time are shown in Figure 2. While no interaction effect was hypothesised, an interaction was observed in the data, $F(1,124)=12.388$, $p=.001$ (See Figure 2). Higher-performing students' calibration levels dropped slightly but not significantly over the semester, $t(70)=1.542$, $p=.128$, while lower ability students' self-efficacy bias decreased significantly over time, meaning that calibration improved over the semester, though students remained overconfident, $t(54)=4.607$, $p<.001$, $d=0.575$, 95%CI[0.194, 0.957]. The main effects are still interpretable in light of the interaction, as both lower- and higher-performing students were more overconfident at mid-semester than at the end of semester, and lower-performing students were more overconfident than higher-performing students on both occasions.

Figure 2

Subject-level Self-efficacy Bias Over Time and by Achievement Level (n=126)
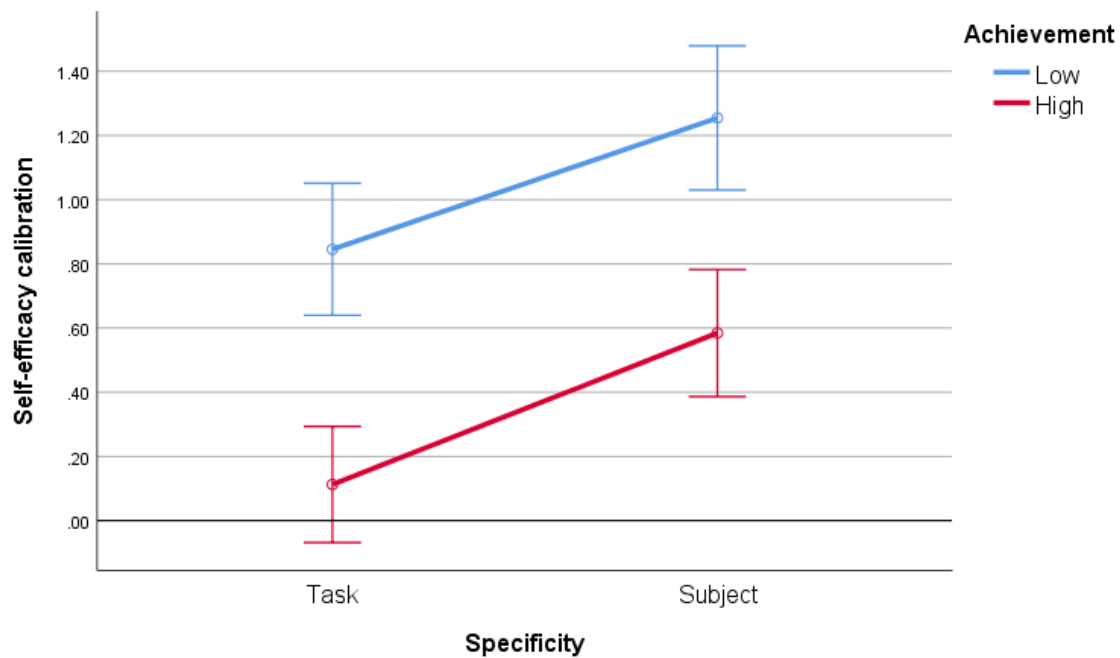

Note: Error bars are 95% confidence intervals. Values on the Y axis indicate the average difference between self-efficacy and academic performance as proportions of a grade band (zero means accurate calibration, positive values mean overconfidence, and negative values mean underconfidence).

Consistent with hypothesis 3, collapsing both time 1 and time 2 measurements, there was a main effect of specificity: students were significantly more overconfident with regard to overall grades (*M*=0.980, 95%CI[0.770, 1.069) than specific task grades (*M*=0.479, 95%CI[0.342, 0.616]), $F(1,124)=74.577$, $p<.001$, *d*=0.819, 95%CI[0.562, 1.077]. Again, there was a main effect of ability: overall self-efficacy bias scores of lower-performing students (*M*=1.050, 95%CI[0.848, 1.252]) exceeded those of higher-performing students (*M*=0.349, 95%CI[0.171, 0.526]), $F(1,124)=26.729$, $p<.001$, *d*=0.926, 95%CI[0.556, 1.296].While the analysis revealed no interaction effect, because both higher- and lower-performing students were more overconfident at subject than task level, $F(1,108)=.509$, $p=0.477$, it is interesting to note how scores relate to zero (calibrated). As shown in Figure 3, the only evidence of self-efficacy calibration was for higher-achieving students at task-level; all other measurements showed consistent overconfidence on average.

Figure 3

Self-efficacy Bias by Specificity and Achievement Level (n=126)

Note: Error bars are 95% confidence intervals. Values on the Y axis indicate the average difference between self-efficacy and academic performance as proportions of a grade band (zero means accurate calibration, positive values mean overconfidence, and negative values mean underconfidence).

Discussion

In this study we considered how first-year undergraduate students' self-efficacy calibration varied across task type and specificity level, over time, and according to students' achievement level. Overconfidence was prevalent, particularly among lower-achieving students and at subject level.

In our sample, no differences in self-efficacy calibration across tasks type, i.e., written assignment and multiple-choice exam, were observed. Previous findings regarding the influences of task type on subjective judgements of capacity are mixed. Inaccuracy in judgements of capacity can result from using cues which are not valid indicators of performance (Koriat, 1997). The lack of any significant difference between the tasks may rest on students using multiple cues with different effects which wash each other out. For example, familiarity with salient features of an exam (i.e., multiple-choice format) (Bol & Hacker, 2001) may lead to overconfidence, while the hard-easy effect (Lichtenstein et al., 1982) or concerns about the "trickiness" of multiple-choice items (Bol et al., 2005) might lead to underconfidence on the same task. Deeper processing of a more complex task may mean that students are better calibrated for a written assessment, but then the flip side of the hard-easy effect would lead to a tendency towards overconfidence. Experience with the multiple-choice format might be traded off against the knowledge that the format necessitates "guesswork", while lesser experience with university written assignments might be traded off against having knowledge of the question in advance (Bol et al., 2005). Ultimately, the null finding here may suggest that task type does not have an impact on self-efficacy bias, or that a difference does exist, but that multiple competing influences on calibration obscured it.

High achievers' subject-level self-efficacy calibration did not change significantly over the semester: they were somewhat overconfident on both occasions. Low achievers were more than twice as overconfident as high achievers at mid-semester, becoming less overconfident by end of semester (though still roughly twice as overconfident as high achievers). Early in students' university careers, subject and course-level self-efficacy judgements are potentially made in the absence of applicable

mastery information, which may result in decreased accuracy (Bandura, 1997; Usher & Pajares, 2008). This is consistent with the findings of Gore Jr (2006), who found that experienced college students' self-efficacy predicted performance better than that of "college-naïve" students. A small amount of research has considered how self-efficacy bias is impacted by interventions designed to reduce over-inflated self-estimations (Zimmerman et al., 2011). The present research suggests that, without any specific intervention, there may be some natural self-correction of overconfidence in lower-performing students over time – perhaps simply as they are exposed to the sobering realities of the grades and feedback they receive during their studies. This contrasts with previous research indicating that overconfidence did not decline over a course of study (Foster et al., 2016), though the difference may be related to the level of specificity, as the Foster et al. study focused on class exams rather than subject-level performance as in the present case.

Both lower- and higher-performing students showed more positively biased self-efficacy scores at subject level than at task level, consistent with other studies showing higher self-evaluations at broader levels of analysis than more specific levels (Ackerman et al., 2002). Deeper processing of task-level cues may be associated with reduced confidence (Gist & Mitchell, 1992), while domain-level judgements may be based on a more general overconfidence bias (Pallier et al., 2002), or desired outcomes (Serra & DeMarree, 2016). Higher overconfidence at subject level suggests that students potentially disregard the trees when judging the forest. This may be a strategy for managing perceived threats to self-worth (Sedikides & Strube, 1997).

A tendency for lower-achieving students to be overconfident was noted consistently across analyses. We interpret this as the social-cognitive theory version of the "unskilled and unaware" effect documented in meta-cognitive judgements (Fritzsche et al., 2018; Kruger & Dunning, 1999). Similar findings are reported for calibration of a range of objective and subjective measures (e.g., Bol et al., 2005; Zimmerman et al., 2011). Our findings suggest that the capacity to perform well and the capacity to judge performance capacity overlap, such that poorer performers are less able to accurately judge their capabilities. On average, higher-achievers were accurately calibrated for specific tasks, and were about half as overconfident as lower-achievers when considering subject performance overall. Higher-achievers' calibration may have been impacted by being more "skilled and aware", or by the factors discussed earlier: defensive pessimism, socially desirable responding, or simply being more humble or cautious (Bol et al., 2005; de Carvalho Filho, 2009; Norem & Cantor, 1986; Schunk & Pajares, 2004; Taylor & Shepperd, 1998).

Limitations and Directions for Future Research

Conclusions drawn from our findings should be tempered by an understanding of the limitations of the present work. In conducting our research in a naturalistic setting, it was not possible to exercise control over all aspects of study. For example, it was not possible to measure self-efficacy for both tasks at the same time as each other; this presents a risk of maturation effects impacting on the later task-level self-efficacy measure. Also, our sample comprised first-year psychology students in a regional Western higher education setting; additional research would be needed to determine the extent to which the present findings apply to other learners.

In order to unravel the complex influences on self-efficacy bias, future research may explore the cues students use to make these types of judgements. It may be possible to experimentally manipulate the types of cues which are salient to students as they make self-efficacy judgements, and compare differences across these. Future research is also warranted to determine whether the changes over time in self-efficacy bias are evident for the same type of specific task. Further research could also shed light on how (mis)calibration relates to student well-being.

Implications

This study's findings add to a body of research which increasingly counsels caution regarding the widespread recommendation to enhance academic self-efficacy in isolation. Advice to bolster self-efficacy is based on the popularly perceived value of having high self-efficacy, which rests on an interpretation of the positive relationship commonly found between self-efficacy and academic performance in the research literature (Honicke & Broadbent, 2016; Richardson et al., 2012) and also in the present study. However, this approach risks two classic blunders in the translation of research: equating correlation with causation, and assuming that group-level statistics apply equally to each individual in the group.

Regarding the first point, the error in this case is to assume that the correlation means "I believe, therefore I achieve". This interpretation persists in spite of research aimed at teasing apart the relationship which has found that the unique impact of self-efficacy on academic outcomes is very small (Talsma, Norris, et al., 2019; Talsma et al., 2018; Valentine et al., 2004).

Regarding the second point, the positive correlation is interpreted to mean that low self-efficacy is always bad and high self-efficacy is always good – regardless of how self-efficacy aligns with the capacities of a given individual. It is assumed, for example, that low achieving students lack the confidence that would help them to perform better, and that this low confidence needs to be remedied. But a correlation coefficient says nothing about how an individual's self-efficacy beliefs relate to that individual's capacity to perform. The positive correlations in our data (and, we speculate, in much existing data) are typified by positive bias: self-efficacy beliefs tend to exceed objective performance outcomes – especially for lower-achievers and especially at subject level. Consider a group of young adult men who are all roughly a few centimetres taller than their fathers – there would be an extremely strong correlation between the heights of the fathers and sons. The correlation tells us that taller fathers have taller sons, but looking at the correlation in isolation potentially obscures the fact that even the shortest son is taller than his own father.

In the same way, while some students' self-efficacy beliefs will be lower than those of some other students, this does not necessarily mean that their beliefs are *too* low relative to their own performance capacity. Self-efficacy theorists encourage a reach that exceeds one's grasp (Pajares, 2006), and attempts to bolster low self-efficacy are likely made with this position in mind. Well-intentioned though this may be, it likely ignores the possibility that even a short reach might already exceed one's grasp, potentially considerably. Self-efficacy theorists agree that gross overestimation of abilities is not adaptive (Bandura, 1995). Other perspectives suggest that even modest overconfidence can result in a perceived lack of any discrepancy between current and desired states, suppressing effortful behaviours or self-regulatory action and potentially leading to poorer performance (Vancouver & Kendall, 2006).

Thus, students who are already overconfident (i.e., most students in our first-year sample, especially at subject-level, and especially low achievers; see Table 2) are unlikely to benefit from approaches aimed at increasing self-efficacy. In this context, it is concerning that approaches to feedback seem to increasingly prioritise affective concerns and the protection of self-beliefs, as this may undermine lower-achieving learners' engagement with accurate diagnostic and "feed-forward" information that could assist in calibrating self-efficacy beliefs and enhancing future performance (Boud, 2015; Dinham, 2010; Hattie & Timperley, 2007; Molloy et al., 2020).

Meanwhile, high achievers who have more realistic task-level self-efficacy beliefs also seem unlikely to benefit from self-efficacy-enhancing interventions, as it is argued that accurately calibrated self-efficacy beliefs benefit self-regulation processes (Bol & Hacker, 2012; Dunlosky &Thiede, 2013; Hattie, 2013). We note that there was a small number of students in our sample who were underconfident – again, these were mostly high achievers considering their self-efficacy for specific tasks. It seems plausible that, for high achieving students who undersell themselves at task level, the sense of discrepancy this engenders is motivating for them (Norem & Cantor, 1986; Vancouver & Kendall, 2006). In our sample, most of the students whose self-efficacy beliefs undersell their abilities are also those least in need of performance improvements. So again, it is unclear whether self-efficacy-enhancing interventions would be of benefit in this case.

There was a very small number of underconfident low achievers in our sample. It would be necessary for teachers to determine on a case-by-case basis whether increased self-efficacy would be beneficial for students in this situation. It is key not only to determine how self-efficacy relates to academic outcomes for individual students, but also to ascertain what impact any miscalibration has on that individual's behaviour. Self-efficacy-enhancing interventions would only be indicated in cases where students were underconfident *and* that underconfidence led to maladaptive learning behaviours. Thus, a one-size fits-all approach to self-efficacy is unlikely to be of benefit to all students.

Conclusion

First-year university students appear to commonly hold self-efficacy beliefs which are miscalibrated with their objective performance outcomes. The nature of this academic self-efficacy bias varies according to characteristics of the students and the academic outcomes in question. Overall, lower-performing students tend to make self-efficacy judgements which considerably exceed their objective performance outcomes, while higher-performing students are more accurate. The strength of this effect depends on the specificity of the academic outcome in question: lower-performing students are overconfident at both task and subject-level, while higher performing students accurately calibrated at task-level, but moderately overconfident at subject-level. Subject-level overconfidence in lower-performing students appears to diminish somewhat over time, while higher-performing students are less overconfident and more stable over time. Task type seems to have little impact on self-efficacy bias, though additional research is necessary.

It is hoped that these findings will contribute to a deeper understanding of the complexities of academic self-efficacy calibration, and prompt further research regarding the potential impact of calibration interventions for different students and contexts. It may be of benefit for both faculty and learners to be aware of the different patterns of self-efficacy accuracy and inaccuracy across judgement specificity and ability levels. Educational counselling and feedback practices which take these differences into account may encourage more adaptive self-efficacy evaluations and contribute to stronger academic performance outcomes, particularly for lower-achieving 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. https://doi.org/10.1016/S0191-8869(01)00174-X

Artino, A. R. (2012). Academic self-efficacy: From educational theory to instructional practice. *Perspectives on Medical Education*, *1*(2), 76-85. https://doi.org/10.1007/s40037-012-0012-5

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

Bandura, A. (1995). *Self-efficacy in changing societies*. Cambridge University Press.

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. https://doi.org/10.1016/j.learninstruc.2009.03.002

Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *The Journal of Experimental Education*, *69*(2), 133-151. https://doi.org/10.1080/00220970109600653

Bol, L., & Hacker, D. J. (2012). Calibration research: Where do we go from here? *Frontiers in Psychology*, *3*, 229: 1-6. https://doi.org/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. https://doi.org/10.3200/JEXE.73.4.269-290

Bong, M. (2001). Role of self-efficacy and task-value in predicting college students' course performance and future enrollment intentions. *Contemporary Educational Psychology*, *26*(4), 553-570. https://doi.org/10.1006/ceps.2000.1048

Boud, D. (2015). Feedback: Ensuring that it leads to enhanced learning. *Clinical Teacher*, *12*, 3-7. https://doi.org/10.1111/tct.12345

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. https://doi.org/10.1080/00207590701436744

Dinham, S. (2010). Self-esteem: Caution – do not over-inflate [online]. *Teacher: The National Education Magazine August 2010*, 6-11. Retrieved from: https://search.informit.com.au/documentSummary;dn=271017548564539;res=IELAPA

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. https://doi.org/10.1016/j.learninstruc.2012.05.002

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. https://doi.org/10.1111/1467-8721.01235

Foster, N. L., Was, C. A., Dunlosky, J., & Isaacson, R. M. (2016). Even after thirteen class exams, students are still overconfident: the role of memory for past exam performance in student predictions. *Metacognition and Learning*, 1-19. https://doi.org/10.1007/s11409-016-9158-6

Fritzsche, E. S., Händel, M., & Kröner, S. (2018). What do second-order judgments tell us about low-performing students' metacognitive awareness? *Metacognition and Learning*, *13*(2), 159-177. https://doi.org/10.1007/s11409-018-9182-9

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. https://doi.org/10.5465/AMR.1992.4279530

Gore Jr, P. A. (2006). Academic self-efficacy as a predictor of college outcomes: Two incremental validity studies. *Journal of Career Assessment*, *14*(1), 92-115. https://doi.org/10.1177/1069072705281367

Hattie, J. (2013). Calibration and confidence: Where to next? *Learning and Instruction*, *24*, 62-66. https://doi.org/10.1016/j.learninstruc.2012.05.009

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81-112. https://doi.org/10.3102/003465430298487

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

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349. https://doi.org/10.1037/0096-3445.126.4.349

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

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art. In *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.

Molloy, E., Ajjawi, R., Bearman, M., Noble, C., Rudland, J., & Ryan, A. (2020). Challenging feedback myths: Values, learner involvement and promoting effects beyond the immediate task. *Medical Education*, *54*(1), 33-39. https://doi.org/10.1111/medu.13802

Norem, J. K., & Cantor, N. (1986). Defensive pessimism: Harnessing anxiety as motivation. *Journal of Personality and Social Psychology*, *51*(6), 1208-1217. https://doi.org/10.1037//0022-3514.51.6.1208

Pajares, F., & Usher, E. L. (2008). Self-efficacy, motivation, and achievement in school from the perspective of reciprocal determinism. In M. L. Maehr, T. C. Urdan, & S. Karabenick (Eds.), *Advances in motivation and achievement* (Vol. 15, pp. 391-423). https://doi.org/10.1016/S0749-7423(08)15012-9

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *Journal of General Psychology*, *129*(3), 257-299. https://doi.org/10.1080/00221300209602099

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. https://doi.org/10.1037/a0026838

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). Information Age Publishing.

Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. *Advances in experimental social psychology*, *29*, 209-269. https://doi.org/10.1016/S0065-2601(08)60018-0

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. https://doi.org/10.3758/s13421-016-0624-9

Talsma, K., Norris, K., & Schüz, B. (2019, July). *Self-efficacy and academic performance: A chicken-and-egg conundrum* [conference paper]. STARS Conference, Melbourne. https://unistars.org/papers/STARS2019/03C.pdf

Talsma, K., Schüz, B., & Norris, K. (2019). Miscalibration of self-efficacy and academic performance: Self-efficacy ≠ self-fulfilling prophecy. *Learning and Individual Differences*, *69*, 182-195. https://doi.org/10.1016/j.lindif.2018.11.002

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. 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. https://doi.org/10.1177/0146167298249001

Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research*, *78*(4), 751-796. https://doi.org/10.3102/0034654308321456

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. https://doi.org/10.1207/s15326985ep3902_3

Vancouver, J. B. (2012). Rhetorical reckoning: A response to Bandura. *Journal of Management*, *38*(2), 465-474. https://doi.org/10.1177/0149206311435951

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. https://doi.org/10.1037/0021-9010.91.5.1146

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. https://doi.org/10.1016/j.sbspro.2010.03.185

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 311-328). Routledge.

Zimmerman, B. J., Moylan, A., Hudesman, J., White, N., & Flugman, B. (2011). Enhancing self-reflection and mathematics achievement of at-risk urban technical college students. *Psychological Test and Assessment Modeling*, *53*(1), 141-160.