



Metacognitive awareness as measured by second-order judgements among university and secondary school students

Marloes L. Nederhand¹  · Huib K. Tabbers¹ · Anique B.H. De Bruin² · Remy M. J. P. Rikers^{1,3}

Received: 3 October 2018 / Accepted: 11 May 2020 / Published online: 28 May 2020
© The Author(s) 2020

Abstract

When compared to high performers, low performers generally have more difficulty to accurately estimate their own performance. This has been explained by low performers being both unskilled and unaware about their performance. However, Miller and Geraci *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(2), 502–506, (2011) found that low performing university students also assigned less confidence to their estimates (i.e., second-order judgments, SOJs), indicating some metacognitive awareness of their poor calibration. The current study examined whether the relationship between calibration accuracy and confidence in performance estimates is more general, and exists irrespective of performance level, not only for university students but also for secondary school students. We asked university students and secondary school students to estimate their exam grade after taking their exam, and to provide a second-order judgement). The results showed that for university students, poor calibration accuracy was indeed accompanied by low confidence scores, independent from performance level. For secondary school students however, calibration accuracy was unrelated to confidence scores, suggesting a less developed metacognitive awareness.

Keywords Calibration accuracy · Overconfidence, second-order judgements · Self-assessment · Unskilled-and-unaware effect · Performance level · Age

✉ Marloes L. Nederhand
m.l.nederhand@essb.eur.nl

¹ Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Woudestein, Mandeville Building, T13-54, P.O. Box 1738, 3000 DR, Rotterdam, The Netherlands

² Department of Educational Development and Research, Maastricht University, Maastricht, The Netherlands

³ The Roosevelt Center for Excellence in Education, University College Roosevelt, Utrecht University, Utrecht, The Netherlands

Introduction

Many students are generally overconfident about their own performance (De Bruin et al. 2017; Dunning et al. 2003; Ehrlinger and Dunning 2003; Kruger and Dunning 1999; Pennycook et al. 2017; Sanchez and Dunning 2018). Not being able to accurately estimate the level of one's performance can have far-reaching implications, as performance estimates influence control decisions made during the learning process (Nelson and Narens 1990). In the case of overconfidence, students may prematurely end studying because of an incorrect conviction they have mastered the materials (Bol et al. 2005; Dunlosky and Rawson 2012). Especially low-performing students seem to suffer from inaccurate self-monitoring, mostly from overconfidence. High performing students are often much better calibrated, although they may show some underconfidence. This difference in over- and underconfidence among low and high performers is dubbed the unskilled-and-unaware effect (Dunning et al. 2003; Kruger and Dunning 1999).

The unskilled and unaware effect sparked discussion in the literature. A first explanation for the effect was that low performers lack metacognitive awareness because of a 'double curse' (Kruger and Dunning 1999). Low performers would lack the skills to conduct a task, and because of this incompetence, they were supposed to also be unaware of the level of their performance. However, other researchers suggested that the effect was simply an artefact of scaling. For example, Krueger and Mueller (2002) and Feld et al. (2017) showed that a large proportion of the unskilled-and-unaware effect could be attributed to regression to the mean. Yet, there remained a proportion of miscalibration and overconfidence of low performers that could not fully be explained by scaling error (Ehrlinger et al. 2008; Feld et al. 2017). It seemed that, albeit less than initially thought, low performers showed less awareness of their own performance than high performers, at least when judging their performance relative to their peers. In a study of Hartwig and Dunlosky (2014), a classic unskilled-and-unaware effect was shown when students compared their performance to a peer group. However, the effect disappeared when students made absolute judgements of their own performance. Hence, the miscalibration of low performing students on percentile judgements seemed to arise from the inability to judge how their reference group would perform, not from a lack of self-awareness (Hartwig and Dunlosky 2014).

Recently, the unskilled-and-unaware effect is also challenged by studies using a different type of metacognitive measurement. Miller and Geraci (2011) argued that focusing solely on performance estimates gives an incomplete image of students' metacognitive awareness. Instead, they argued that calibration research should focus on both functional and subjective confidence. Miller and Geraci defined functional confidence as students' performance estimate, and subjective confidence as the confidence judgement assigned to a given performance estimate. Such subjective confidence judgements are studied by asking students to provide a second-order judgement (SOJ), i.e., rating the confidence they had in their estimate (Dunlosky et al. 2005). Using confidence judgements in calibration research is not new. Many studies on metacognitive awareness use confidence judgements (e.g., Lichtenstein and Fischhoff 1977; Siedlecka et al. 2016, 2019). For example, participants were asked to select a certain answer to a question and to rate how confident they were that their selected answer was indeed accurate (e.g., Dinsmore and Parkinson 2013; Siedlecka et al. 2016). This is, however, different from SOJs. Instead of judging how confident they are in their answer, students judge how confident they are in their estimate when providing a SOJ. Although Dunlosky et al. (2005) pointed to the value of SOJ's to gain more insight in metacognitive awareness, it is relatively recent that

SOJs really came at the center of attention (Miller and Geraci 2011; Händel and Fritzsche 2013, 2016). The reason for this is that SOJs may shed more light on the question whether low performers are indeed unaware of their poor performance, or whether they do show some metacognitive awareness by providing low confidence scores to their performance estimate.

In their study, Miller and Geraci asked university students not only to estimate their grade before taking an exam, but also to provide a 'second-order judgement' (SOJ; see Dunlosky et al. 2005): rating the confidence they had in their estimate. In line with previous research, Miller and Geraci found that low performers showed a larger mismatch between their estimated grade and actual grade than high performers. At the same time however, low performers reported less confidence in the accuracy of their estimates than high performers. According to Miller and Geraci, low performers thus seemed to have a certain level of metacognitive awareness. Following up on the research of Miller and Geraci (2011), Händel and Fritzsche (2016) investigated the alignment between SOJs and calibration accuracy on item-by-item judgements and global judgements (i.e., judgements regarding performance on an entire task). Händel and Fritzsche also found a positive relation between performance level and SOJs. Low performers, who were found to be poorly calibrated, were again less confident in their performance estimates than high performers.

Hence, the studies of Miller and Geraci (2011) and Händel and Fritzsche (2016) showed that low performers seem to be aware of their poor calibration accuracy. This finding has implications for the explanation of the unskilled-and-unaware effect, because, as also stated by Hartwig and Dunlosky (2014), qualifying low performers as fully unaware of their performance seems to be incorrect. However, in a follow-up study, Fritzsche et al. (2018) reanalyzed the data from Händel and Fritzsche (2016), and found that low performing students seemed to have a general tendency to provide low confidence judgements. This finding raises the issue whether it is performance level that determines whether students provide high or low subjective confidence scores, or whether it is actual metacognitive awareness.

If SOJs could be used to measure metacognitive awareness, we would expect subjective confidence to be directly related to calibration accuracy. This means that students high on miscalibration have little confidence in their judgements, and students low on miscalibration show high confidence, irrespective of students' performance level. In that case, students would be metacognitively aware, with confidence judgements aligned to their calibration accuracy.

However, although these prior studies all aim to examine whether people are able to match their confidence judgements to their calibration accuracy, none of the previous studies actually examined the direct relation between calibration accuracy on the one hand, and SOJs on the other, independent from performance level. Hence, the question remains whether lower or higher SOJs are caused by differences in performance, or in students' ability to calibrate accurately.

Another question that remains unanswered is whether findings on SOJs generalize to different age and school groups. While the studies discussed in the introduction have been exclusively conducted using university students (e.g., Fritzsche et al. 2018; Händel and Fritzsche 2016; Miller and Geraci 2011), unskilled-and-unaware effects have been found for younger students as well (e.g., Finn and Metcalfe 2014; Labuhn et al. 2010). At the same time, however, students and younger children are not completely unaware of their performance. Although research shows a clear age trend in metacognitive awareness (Roebbers and Schneider 2005; Schneider 2008), from the age of 8, children show some metacognitive awareness on simple tasks (De Bruin et al. 2011; Lyons and Ghetti 2011; Roebbers 2017, see also Schneider 2008). However, when tasks are more complicated and metacognitive skills are more

advanced, for example because students have to show transfer skills or when they have to recognize whether they guessed or not, age trends reappear (Hacker 1997; Lockl and Schneider 2002; Roebbers 2014; Van der Stel and Veenman 2010; Weil et al. 2013). In such situations, the basic metacognitive skills need to be fine-tuned which continues to happen in adolescence (Paulus et al. 2014), after which it plateaus in adulthood (Roebbers 2014; Weil et al. 2013). Providing SOJs that align to the accuracy of one's estimate (i.e., low confidence for inaccurate estimate, high confidence for accurate estimate) is a more complex form of metacognitive awareness, especially when the tasks on which such judgements have to be made consist of problem-solving. It is therefore the question whether findings on university students' subjective confidence can be generalized to younger secondary school students. Alternatively, secondary school students may be less metacognitively aware, and show less alignment between their subjective confidence and actual calibration accuracy.

Present study

The current study thus aimed to answer two research questions: (1) are SOJs related to calibration accuracy, independent from performance level?; and (2) does the negative relation between SOJs and miscalibration hold for both university students and secondary school students?

To investigate the relation between subjective confidence and calibration accuracy, we conducted two studies in which students were asked to estimate their exam grade and to provide a confidence score of their estimate (subjective confidence as measured by a SOJ) after taking their exam. In the first study, we examined the alignment between SOJs and calibration accuracy among university students who took an exam in educational psychology (cf. Miller and Geraci 2011). The second study was conducted among secondary school students who took exams in three different subjects: French, German, and Mathematics.¹ For each exam, we investigated the alignment between calibration accuracy and SOJs.

Based on Miller and Geraci, we expected a significant negative relation between SOJs and calibration accuracy. That is, higher miscalibration would be associated to lower confidence. Furthermore, because developmental differences might affect metacognitive awareness, we expected that secondary school students would be less metacognitively aware and thus the relation between calibration accuracy and SOJs would be less strong in this sample.

Finally, we checked whether we could replicate the association between performance level and SOJ that Miller and Geraci (2011) found. Similar to Miller and Geraci, we divided our students into different quartiles based on performance level, to examine the differences in SOJs among high and low performers. Based on the assumption that low performers indeed calibrate worse than high performers (cf. the unskilled-and-unaware effect), we expected low performers to also provide lower confidence scores.

¹ The secondary school students participated in an intervention study in which we trained their calibration accuracy over time. The data reported in this study consist of their pre-measurement scores, before they took part in the intervention.

Method

Participants

Two-hundred-and-ninety-four second-year psychology students and 388 secondary school students from the same urban area in the Netherlands were recruited for this study. Confidentiality of the performance estimates was ensured to all students, and they all provided informed consent that their estimates and grades could be used for this study. Ethical approval was obtained for this study by the Ethical Committee of the Department of Psychology, Education, and Child Studies at our university. Two-hundred-and-fifty-one university students provided informed consent (85.37% response rate). The secondary school students provided informed consent together with their parents, leading to a final sample of 302 students (77.84% response rate).

Materials

Exams We used students' regular exams to measure calibration accuracy. The university students had an end-of-term exam in educational psychology, comparable to the exams used in most research on calibration accuracy with university students (Bol and Hacker 2001; Hacker et al. 2000; Miller and Geraci 2011; Nietfeld et al. 2006). During their academic year in which they took 8 exams, they could resit a maximum of 2 exams. The secondary school students had regular end of term exams in French, German, and Math. None of the students were allowed to resit unless they were sick when the exam was administered.

Procedure

The procedure was the same for all students. After taking their exam, students received a form on which they estimated the grade they thought they would obtain on a 10-point scale with decimal points (i.e., most common scoring scale in the Netherlands with 10 representing the highest possible score and 1 the lowest). Students also gave a SOJ by indicating the confidence they had in their performance estimate on a five-point scale, ranging from not confident to highly confident (cf. Händel and Fritzsche 2016; Miller and Geraci 2011).

Analyses

Calculating measurements

Calibration accuracy was calculated by taking the absolute difference between estimated exam grade and actual exam grade (Dunlosky and Thiede 2013; Schraw 2009). This means that the higher the difference between the estimated and actual grade, the higher the miscalibration. We also calculated bias scores to investigate the direction of the miscalibration (i.e., over- or underconfidence). Bias scores consisted of the difference between estimated grade and actual grade (Dunlosky and Thiede 2013; Schraw 2009).

Statistical analyses

Our research question was whether subjective confidence was aligned with calibration accuracy scores. To test this question, we investigated Pearson correlations between SOJs and calibration accuracy for each exam. As a check, we also investigated whether we could replicate the findings of Miller and Geraci (2011). To do so, we divided the university students into four quartiles based on their exam performance. We tested whether SOJs differed between low performers and high performers (cf. Miller and Geraci 2011) using an ANOVA with SOJ as dependent variable and performance level as independent variable.

Results

Table 1 shows the descriptives of the variables under study: Bias scores, calibration accuracy, SOJs, and final exam grades. In the following sections, our hypotheses on calibration accuracy are tested. In all our analyses, a significance level of .05 was used.

As a first check, we examined whether we could replicate the findings of Miller and Geraci (2011), by investigating whether possible differences in subjective confidence within our samples could be explained by performance level differences. Table 2 presents the means and standard deviations of the four different performance level groups on calibration accuracy, bias, SOJs, and performance, both for university students on their Educational Psychology exam and for secondary school students for their French, Math and German exam. We performed an ANOVA with SOJ as dependent variable and performance level (i.e., the performance quartiles) as independent variable (cf. Miller and Geraci 2011). Results were not significant: subjective confidence ratings did not differ between the performance quartiles for the university students, $F(3, 251) = 0.07$, $MSE = 0.05$, $p = .978$, $\eta_p^2 = .001$, nor for the secondary school students on the courses French, $F(3, 291) = 1.11$, $MSE = 0.77$, $p = .345$, $\eta_p^2 = .011$, Math, $F(3, 287) = 0.42$, $MSE = 0.82$, $p = .741$, $\eta_p^2 = .004$, and German $F(3, 298) = 0.44$, $MSE = 0.56$, $p = .725$, $\eta_p^2 = .004$. This indicates that, contrary to Miller and Geraci's findings, subjective confidence scores in our samples did not differ between high and low performing students.

Relation between calibration accuracy and second-order judgements among university students

More importantly, however, we examined the direct relation between calibration accuracy and SOJs, irrespective of the level of performance. Results among university students showed a significant negative correlation between subjective confidence (SOJ) and calibration accuracy $r = -.197$, $p = .002$, confirming our hypothesis. Figure 1 shows that less confident students had higher miscalibration, and vice versa: confident students were shown to be better calibrated.

Relation between calibration accuracy and second-order judgements among secondary school students

Thirdly, we tested whether the direct relation between calibration accuracy and SOJs also holds among secondary school students. In contrast to university students, results showed no significant correlation between SOJs and calibration accuracy among secondary school

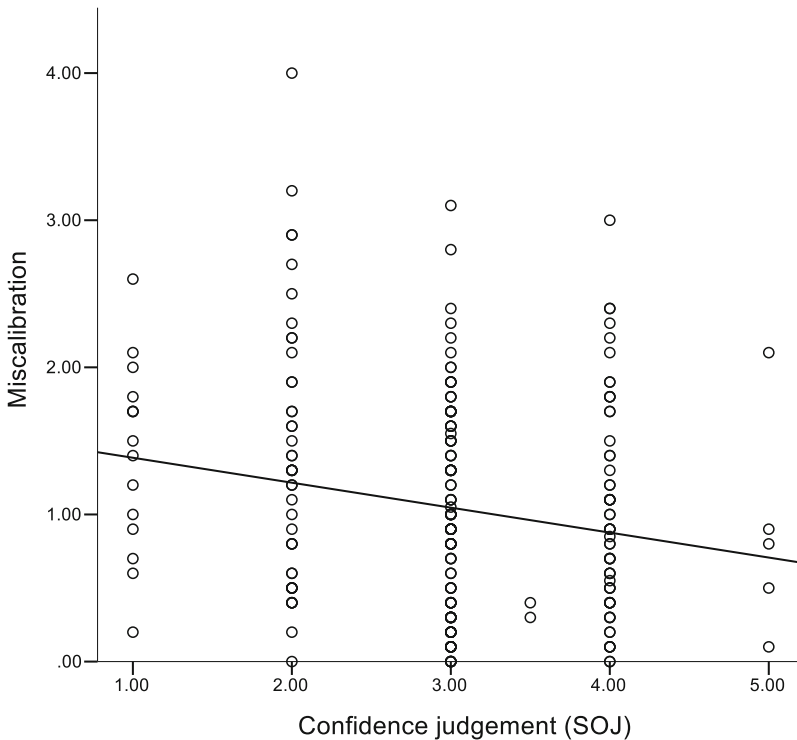


Fig. 1 The negative relation between second-order judgements and miscalibration among university students. Low confidence judgements are related to high miscalibration, whereas high confidence judgements are related to lower miscalibration

students for French $r = -.029$, $p = .654$; Math $r = .036$, $p = .539$; and German $r = .059$, $p = .304$. Hence, confirming our expectation, secondary school students' calibration accuracy and SOJs did not seem to be directly related: more miscalibration was not accompanied by lower confidence scores.

Discussion

The current study focused on the relation between calibration accuracy and subjective confidence (as measured by SOJs) among university and secondary school students. We tested the hypothesis that students, regardless of their performance level, are metacognitively aware of their calibration accuracy. To do so, we examined the correlations between subjective confidence and miscalibration. Furthermore, we tested whether a relation between subjective confidence and miscalibration could be found both among university students and secondary school students.

Calibration accuracy and SOJs among university students

In support of our expectations, university students' subjective confidence is directly related to calibration accuracy, meaning that their confidence judgements are somewhat aligned to the

actual quality of their performance estimates. Our results extend previous research findings on subjective confidence (Händel and Fritzsche 2016; Miller and Geraci 2011). While the studies of Miller and Geraci (2011) and Händel and Fritzsche (2016) were initially aimed at investigating subjective confidence among different performance level groups only, our study shows that their findings generalize to students with poor calibration accuracy in general. That is, the current study shows that students assign less confidence to their judgements when they are less well calibrated. Hence, regardless of performance level, students show some awareness of their calibration accuracy by providing high or low SOJs. This indicates that metacognitive awareness as measured by SOJs does not (only) seem to depend on performance level, as can be concluded from the findings from Miller and Geraci and Händel and Fritzsche, but rather seems to hinge on calibration accuracy.

Supporting the reasoning above that SOJs are more dependent on calibration accuracy than on performance level per se, we do not find any differences in SOJs between different performance quartiles in our sample. This is in contrast to prior findings by Miller and Geraci (2011) and Händel and Fritzsche (2013, 2016). A possible explanation may be that in our study, the relation between performance level and calibration accuracy is different from previous studies. Although we find that on average, high performers are underconfident while low performers are (slightly) overconfident, which is conform a typical unskilled-and-unaware effect, our results show that overall, underconfidence is a larger 'problem' than overconfidence: 3 out of the 4 performance quartile groups show underconfidence. Furthermore, unusual for calibration research, pairwise comparisons even show that the 25% best performing students calibrate slightly worse ($M=1.70$, $SD=0.80$) than the 25% lowest performing students ($M=0.79$, $SD=0.54$), $p < .001$. The current study could therefore not support an unskilled-and-unaware effect. In itself, this supports the notion that having a poor performance does not necessarily make you unaware of this performance: those who showed the poorest performance, did not show the poorest calibration accuracy.

It is yet unclear why our results differ from previous studies that show clear unskilled-and-unaware effects. A possible reason for our findings could be that in many prior studies, predictions were examined (e.g., Bol et al. 2005; Foster et al. 2017; Miller and Geraci 2011) whereas in our study, postdictions were examined. When postdicting performance, students can make use of both theory-based cues (prior performance on exams) and experience-based cues (how easy the student experienced the exam, see also Händel and Bukowski 2019). When predicting their performance, students do not have any experience-based cues and can only rely on theory-based cues. Consequently, findings show that the relation between confidence judgements and actual performance outcomes is lower for predictions than for postdictions (Siedlecka et al. 2016, 2019). It is possible that when postdicting their performance, lower performing students have less difficulty judging their performance because they have access to experience cues (i.e., they found the exam difficult). As a consequence, the difference between groups of students and how well they can judge their performance may disappear. Yet, there are a few studies that examined the unskilled-and-unaware effect with postdictions as well, and they still found an unskilled-and-unaware effect (Hacker et al. 2000; Shake and Shulley 2014). Furthermore, even when cues are available, low performing students usually have more difficulty using valid cues when judging their performance (Gutierrez de Blume et al. 2017; Thiede et al. 2010). So, although the differences between postdictions

and predictions may explain some of the current findings, the difference does not appear to give a definitive answer to the question why students did not show an unskilled-and-unaware effect. Future research that more directly measures students' cue use during pre- and postdictions can shed light on causes for the unskilled-unaware effect.

The purpose of the current study was to examine whether there was a relation between confidence and calibration accuracy. Our data confirm our hypothesis that confidence level is directly related to calibration accuracy. This further strengthens the generalizability of the idea that university students base their confidence judgements on the accuracy of their performance estimate. Again, the reason for doing so likely originated from their cue-use (Händel and Bukowski 2019; Koriat 1997). For example, when judging their exam performance and noting that they had found the exam very difficult, this experience can be taken into account when providing a SOJ. Furthermore, when students have the feeling that they miss certain cues to make an accurate judgement, they can also assign little confidence to their estimate. In other words, students can be aware that their estimate may be incorrect and hence assign lower confidence to them. This can also work vice versa. When students recognize some important cues (i.e., recognize certain tasks from their practice) they used when judging their performance, their confidence in this judgement is likely to be higher.

Calibration accuracy and SOJs among secondary school students

Whereas university students' subjective confidence is related to calibration accuracy, the secondary school students' confidence judgements do not relate to their actual calibration accuracy. This finding seems to be rather robust, as we do not find a relationship in either of the three exams (French, German, and Math). So it seems that our findings with university students do not straightforwardly generalize to secondary school students. Furthermore, this finding supports the notion that metacognitive awareness in secondary school is still under development, especially for more complex tasks and metacognitive judgements (Hacker 1997; Lockl and Schneider 2002; Roebers 2014; Van der Stel and Veenman 2010; Weil et al. 2013). Whereas university students seem to have some metacognitive awareness as they assign little confidence to incorrect estimates, secondary school students seem less able to recognize poor calibration accuracy. This also has an interesting practical implication: when the aim is to improve students' calibration accuracy, secondary school students may first need to be stimulated to reflect on the quality of their performance estimates. Helping secondary school students to become more aware of the poor alignment between their confidence scores and their calibration accuracy, can help them to improve their overall performance monitoring. Interventions that aim to enhance calibration accuracy in secondary school may therefore focus not only on functional confidence, but on subjective confidence as well.

Limitations and future directions

Our study provides new insight in the relation between calibration accuracy and SOJs, both among university and secondary school students. However, there are several limitations to our study as well. Although we find differences in the relation between confidence judgements and calibration accuracy, direct comparisons between the two

samples are confounded by many variables. For example, in university, students' exam grades decide whether they will pass or fail the course and can thus be considered a high-stake exam. In secondary school, the grades add up to one another to form a grand mean at the end of the year. At the same time, however, university students can do a resit whereas secondary school students cannot. It is possible that such differences influence the judgements students make, and hence, their subjective confidence scores. In the current study, we cannot rule out this possibility because we used the exams naturally present in the classrooms. To get more insight in this topic and to enable better comparison between secondary and university students, future research should further examine the role of the task and its weighing on the relation between SOJs and calibration accuracy.

That we do not find significant differences in subjective confidence scores among students of different performance quartiles also contradicts the recent finding that low performers have a general tendency to provide lower confidence judgements than high performers (Fritzsche et al. 2018). It is important to note that our study differs from the study of Fritzsche et al. (2018) in several aspects. Perhaps most importantly, we ask students their exam grade while Fritzsche et al. asked students to estimate their performance while solving multiple choice questions. Their calibration accuracy was therefore based on a dichotomous scale (correct-incorrect), whereas in our study, a continuous scale was used. To facilitate comparison between studies, future research should focus on the question whether differences in scaling impact confidence judgements and calibration accuracy.

Conclusion

To conclude, this study shows that university students' calibration scores are related to their second-order judgements: independent from performance level, students with poor calibration accuracy show less confidence in their estimates and vice versa. Secondary school students, however, provide confidence judgements that are not related to their actual calibration accuracy. Hence, to further understand metacognitive awareness, we encourage future research to take into account measures of both functional and subjective confidence (SOJs). Furthermore, as younger students show less metacognitive awareness, they may benefit less from an intervention that only targets their functional overconfidence, but might instead flourish under an intervention that also aims to align their subjective confidence.

Funding information This research was funded by a Research Excellence Initiative grant from Erasmus University Rotterdam.

Compliance with ethical standards

Conflict of interest No potential conflict of interest was reported by the authors.

Appendix

Table 1 Descriptives of the university and secondary school students

	N	M	[min, max]	SD
University students				
Bias	251	-0.67	[-4.00, 2.10]	1.10
Calibration	251	1.05	[0.00, 4.00]	0.74
SOJ	251	2.98	[1.00, 5.00]	0.86
Final grade	251	7.00	[1.90, 9.70]	1.32
Secondary school students ¹				
French				
Bias	295	-0.19	[-4.70, 3.10]	1.26
Calibration	295	1.02	[0.00, 4.70]	0.75
SOJ	295	3.23	[1.00, 5.00]	0.83
Final grade	295	6.67	[3.00, 9.70]	1.45
Math				
Bias	291	-0.27	[-5.10, 4.80]	1.57
Calibration	291	1.24	[0.00, 5.10]	1.00
SOJ	291	3.34	[1.00, 5.00]	0.90
Final grade	291	6.86	[3.00, 10.00]	1.47
German				
Bias	302	0.58	[-2.10, 3.20]	1.18
Calibration	302	1.07	[0.00, 3.20]	0.76
SOJ	302	3.32	[1.00, 5.00]	0.75
Final grade	302	6.32	[3.00, 9.40]	1.36

This Table presents the bias scores, on a scale of -10 to 10; calibration scores, on a scale of 0 to 10; SOJ scores, on a scale of 1 to 5; and performance scores, on a scale of 1 to 10, in which a score of ≥ 5.5 indicates that students passed the exam

¹ Secondary school students had to take three exams (French, German, Math), but due to sickness or personal reasons, students sometimes failed to be present at all three exams. The *N* therefore varies between subjects

Table 2 Performance quartiles among the university and secondary school students

Quartiles	<i>N</i>	Bias <i>M (SD)</i>	Calibration <i>M (SD)</i>	SOJ <i>M (SD)</i>	Estimate <i>M (SD)</i>	Performance <i>M (SD)</i>
University students						
1	68	0.37 (0.89)	0.79 (0.54)	2.96 (1.00)	5.74 (1.02)	5.37 (0.84)
2	63	-0.53 (0.81)	0.73 (0.62)	2.97 (0.90)	6.16 (0.79)	6.69 (0.26)
3	66	-1.03 (0.72)	1.09 (0.64)	3.02 (0.83)	6.58 (0.73)	7.61 (0.25)
4	54	-1.70 (0.80)	1.70 (0.80)	3.00 (0.67)	7.00 (0.73)	8.69 (0.46)
Secondary school students						
French						
1	82	0.63 (1.22)	1.14 (0.76)	3.17 (0.94)	5.53 (1.23)	4.90 (0.73)
2	70	0.04 (1.22)	0.97 (0.73)	3.17 (0.92)	6.26 (1.22)	6.22 (0.24)
3	70	-0.52 (0.98)	0.84 (0.71)	3.20 (0.71)	6.74 (1.03)	7.26 (0.33)
4	73	-1.00 (0.91)	1.11 (0.78)	3.38 (0.72)	7.53 (1.02)	8.53 (0.48)
Math						
1	74	0.80 (1.44)	1.36 (0.93)	3.42 (1.03)	5.72 (1.51)	4.92 (0.84)
2	82	-0.05 (1.40)	1.11 (0.84)	3.24 (0.92)	6.50 (1.43)	6.55 (0.31)
3	51	-0.70 (1.27)	1.07 (0.96)	3.39 (0.80)	6.70 (1.31)	7.39 (0.16)
4	84	-1.16 (1.39)	1.36 (1.19)	3.35 (0.81)	7.39 (1.28)	8.55 (0.56)
German						
1	77	1.31 (1.04)	1.45 (0.83)	3.19 (0.83)	6.08 (1.23)	4.77 (0.51)
2	74	1.20 (0.92)	1.29 (0.79)	3.39 (0.70)	6.77 (0.92)	5.57 (0.26)
3	81	0.26 (0.95)	0.78 (0.59)	3.31 (0.77)	7.10 (0.85)	6.85 (0.38)
4	70	-0.52 (0.76)	0.75 (0.53)	3.41 (0.67)	7.70 (0.92)	8.22 (0.51)

This Table presents the bias scores, on a scale of -10 to 10; calibration scores, on a scale of 0 to 10; SOJ scores, on a scale of 1 to 5; both performance estimates and performance scores on a scale of 1 to 10, in which a score of ≥ 5.5 indicates that students passed the exam

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

References

- Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *Journal of Experimental Education*, 69(2), 133–151. <https://doi.org/10.1080/00220970109600653>.
- 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. *Journal of Experimental Education*, 73(4), 269–290. <https://doi.org/10.3200/JEXE.73.4.269-290>.
- De Bruin, A. B., Thiede, K. W., Camp, G., & Redford, J. (2011). Generating keywords improves metacomprehension and self-regulation in elementary and middle school children. *Journal of Experimental Child Psychology*, 109(3), 294–310. <https://doi.org/10.1016/j.jecp.2011.02.005>.

- De Bruin, A. B. H., Kok, E., Lobbstaal, J., & De Grip, A. (2017). The impact of an online tool for monitoring and regulating learning at university: Overconfidence, learning strategy, and personality. *Metacognition and Learning, 12*(1), 21–43. <https://doi.org/10.1007/s11409-016-9159-5>.
- Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24*, 4–14. <https://doi.org/10.1016/j.learninstruc.2012.06.001>.
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280. <https://doi.org/10.1016/j.learninstruc.2011.08.003>.
- Dunlosky, J., & Thiede, K. W. (2013). Metamemory. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 283–298). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195376746.013.0019>.
- Dunlosky, J., Serra, M. J., Matvey, G., & Rawson, K. A. (2005). Second-order judgments about judgments of learning. *Journal of General Psychology, 132*(4), 335–346. <https://doi.org/10.3200/GENP.132.4.335-346>.
- 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>.
- Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology, 84*(1), 5–17. <https://doi.org/10.1037/0022-3514.84.1.5>.
- Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) selfinsight among the incompetent. *Organizational Behavior and Human Decision Processes, 105*(1), 98–121. <https://doi.org/10.1016/j.obhdp.2007.05.002>.
- Feld, J., Sauermann, J., & De Grip, A. (2017). Estimating the relationship between skill and overconfidence. *Journal of Behavioral and Experimental Economics, 68*, 18–24. <https://doi.org/10.1016/j.socec.2017.03.002>.
- Finn, B., & Metcalfe, J. (2014). Overconfidence in children's multi-trial judgments of learning. *Learning and Instruction, 32*, 1–9. <https://doi.org/10.1016/j.learninstruc.2014.01.001>.
- 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*, 159–177. <https://doi.org/10.1007/s11409-018-9182-9>.
- Foster, N. L., Was, C. A., Dunlosky, J., & Isaacson, R. M. (2017). Even after thirteen class exams, students are still overconfident: the role of memory for past exam performance in student predictions. *Metacognition and Learning, 12*(1), 1–19.
- Gutierrez de Blume, A. P., Wells, P., Davis, A. C., & Parker, J. (2017). “You can sort of feel it”: Exploring metacognition and the feeling of knowing among undergraduate students. *The Qualitative Report, 22*(7), 2017–2032 Retrieved from <http://nsuworks.nova.edu/tqr>.
- Hacker, D. J. (1997). Comprehension monitoring of written discourse across early-to-middle adolescence. *Reading & Writing, 9*, 207–240. <https://doi.org/10.1023/A:1007989901667>.
- 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. <https://doi.org/10.1037/0022-0663.92.1.160>.
- Hartwig, M. K., & Dunlosky, J. (2014). The contribution of judgment scale to the unskilled-and-unaware phenomenon: How evaluating others can exaggerate over- (and under-) confidence. *Memory and Cognition, 42*, 164–173. <https://doi.org/10.3758/s13421-013-0351-4>.
- Händel, M., & Fritzsche, E. S. (2013). Students' confidence in their performance judgements: a comparison of different response scales. *Educational Psychology, 35*(3), 377–395. <https://doi.org/10.1080/01443410.2014.895295>.
- Händel, M., & Fritzsche, E. S. (2016). Unskilled but subjectively aware: Metacognitive monitoring ability and respective awareness in low-performing students. *Memory and Cognition, 44*(2), 229–241. <https://doi.org/10.3758/s13421-015-0552-0>.
- Händel, M., & Bukowski, A. K. (2019). The Gap Between Desired and Expected Performance as Predictor for Judgment Confidence. *Journal of Applied Research in Memory and Cognition, 8*(3), 347–354.
- Thiede, K. W., Griffin, T. D., Mary, J. W., & Anderson, C. M. (2010). Poor Metacomprehension Accuracy as a Result of Inappropriate Cue Use. *Discourse Processes, 47*(4), 331–362.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>.
- 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. <https://doi.org/10.1037/0022-3514.82.2.180>.

- 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), 349–370.
- Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning*, *5*(2), 173–194. <https://doi.org/10.1007/s11409-010-9056-2>.
- Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Decision Processes*, *20*, 159–183.
- Lockl, K., & Schneider, W. (2002). Developmental trends in children's feeling-of-knowing judgements. *International Journal of Behavioral Development*, *26*(4), 327–333. <https://doi.org/10.1080/01650250143000210>.
- Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development*, *82*(6), 1778–1787. <https://doi.org/10.1111/j.1467-8624.2011.01649.x>.
- Miller, T. M., & Geraci, L. (2011). Unskilled but aware: Reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 502–506. <https://doi.org/10.1037/a0021802>.
- Nelson, T., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation*, *26*, 125–173. [https://doi.org/10.1016/S0079-7421\(08\)60053-5](https://doi.org/10.1016/S0079-7421(08)60053-5).
- Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning*, *1*(2), 159–179. <https://doi.org/10.1007/s10409-006-9595-6>.
- Paulus, M., Tsalas, N., Proust, J., & Sodian, B. (2014). Metacognitive monitoring of oneself and others: Developmental changes during childhood and adolescence. *Journal of Experimental Child Psychology*, *122*(1), 153–165. <https://doi.org/10.1016/j.jecp.2013.12.011>.
- Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic Bulletin and Review*, *24*, 1774–1784. <https://doi.org/10.3758/s13423-017-1242-7>.
- Roebers, C. M. (2014). Children's deliberate memory development: The contribution of strategies and metacognitive processes. In P. J. Bauer & R. Fivush (Eds.), *The Wiley handbook on the development of children's memory* (Vol. 2, pp. 865–894). West Sussex, UK: Wiley-Blackwell.
- Roebers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review*, *45*, 31–51. <https://doi.org/10.1016/j.dr.2017.04.001>.
- Roebers, C. M., & Schneider, W. (2005). The strategic regulation of children's memory performance and suggestibility. *Journal of Experimental Child Psychology*, *91*(1), 24–44. <https://doi.org/10.1016/j.jecp.2005.01.001>.
- Sanchez, C., & Dunning, D. (2018). Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology*, *114*(1), 10–28. <https://doi.org/10.1037/pspa000102>.
- Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education*, *2*(3), 114–121. <https://doi.org/10.1111/j.1751-228X.2008.00041.x>.
- Schraw, G. (2009). Measuring metacognitive judgements. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 439–462). New York, NY: Routledge. <https://doi.org/10.4324/9780203876428.ch21>.
- Shake, M. C., & Shulley, L. J. (2014). Differences between functional and subjective overconfidence in postdiction judgments of test performance. *Electronic Journal of Research in Educational Psychology*, *12*, 263–282. <https://doi.org/10.14204/ejrep.33.14005>.
- Siedlecka, M., Paulewicz, B., & Wierchoń, M. (2016). But I was so sure! Metacognitive judgments are less accurate given prospectively than retrospectively. *Frontiers in Psychology*, *7*, 218. <https://doi.org/10.3389/fpsyg.2016.00218>.
- Siedlecka, M., Skóra, Z., Paulewicz, B., Fijałkowska, S., Timmermans, B., & Wierchoń, M. (2019). Responses improve the accuracy of confidence judgements in memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 712–723. <https://doi.org/10.1037/xlm0000608>.
- Van der Stel, M., & Veenman, M. V. J. (2010). Development of metacognitive skillfulness: A longitudinal study. *Learning and Individual Differences*, *20*(3), 220–224. <https://doi.org/10.1016/j.lindif.2009.11.005>.
- Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., Dolan, R. J., & Blakemore, S. J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition*, *22*(1), 264–271. <https://doi.org/10.1016/j.concog.2013.01.004>.