Title
An Alignment of Standards Enhances Metacognitive Judgment Accuracy in Explanatory Knowledge Tasks with Internet Search

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Mattes, Björn
Pieschl, Stephanie

Publication Date
2022

Peer reviewed

An Alignment of Standards Enhances Metacognitive Judgment Accuracy in Explanatory Knowledge Tasks with Internet Search

Björn Mattes (bjoern.mattes@tu-darmstadt.de)
Pädagogische Psychologie, Alexanderstraße 10, 64283
Darmstadt, Germany

Stephanie Pieschl (stephanie.pieschl@tu-darmstadt.de)
Pädagogische Psychologie, Alexanderstraße 10, 64283
Darmstadt, Germany

Abstract

Previous research indicates that using the internet in knowledge related tasks increases overestimation. We attempted to replicate this finding and extended previous research by explicitly manipulating the standards that participants used for the explanatory knowledge task in order to reduce the metacognitive bias. We conducted a 2x2 within-subject experiment with N = 166 participants. Replicating previous findings, the results show significantly more overestimation in Internet than in No-Internet conditions. However, with an alignment to external standards participants elicited more accurate metacognitive judgments. We conclude that explicit standards may be an important factor in knowledge-related activities involving the internet because of their effect on metacognitive judgments. On a theoretical level, this has implications for determining the basis of overestimation in knowledge tasks with the internet. On a practical level, providing external standards could be a feasible aid for buffering against this bias, for example in the educational context.

Keywords: Metacognition; Internet; Explanatory Knowledge

Introduction

The internet has profound effects on human attention, memory, navigation, and social cognition (Firth et al., 2019; Loh & Kanai, 2016; Marsh & Rajaram, 2019). Nowadays, most of our interactions with information are directly or indirectly centred around the internet and the remarkable shift towards the almost instantaneous access to information has significantly shaped how information is obtained and processed (Greene, Cartiff & Duke, 2018; Ward, 2013). The way information is retrieved with the help of the internet, for example, has changed how such information is remembered (Sparrow, 2011). Importantly, recent studies have shown that the internet has also changed the way we think about our own knowledge and knowledge in general.

In a series of experiments, Fisher, Goddu and Keil (2015) observed that searching online regarding explanations for a first set of explanatory knowledge questions lead to higher self-assessed knowledge in a second unrelated set of explanatory knowledge questions. The authors concluded that the internet might inflate estimates of internal knowledge and that people falsely think they have more knowledge “in the head” than they actually have. While these studies directly compared conditions of internet and no-internet use, they did not, however, capture performance and were therefore unable to show that participants actually overestimated their abilities in answering the explanatory knowledge questions.

Extending the paradigm of Fischer et al. (2015), Pieschl (2021) collected and evaluated participants’ actual answers of the same explanatory knowledge questions. Thus, participants’ metacognitive confidence judgments could be directly related to their corresponding performance. Results show that participants using the internet were indeed biased towards overconfidence. Importantly, this overconfidence went above and beyond the bias of participants answering explanatory knowledge questions without the internet. Additionally, the detected overestimation bias extended from predictive to postdictive metacognitive judgments, with a more pronounced predictive overestimation bias.

Using a different paradigm, with different types of questions, Dunn et al. (2021) studied the influence of internet usage on answering fact-based general knowledge questions. Participants of this experiment were also significantly more accurate in judging their performance without the internet. Furthermore, the internet led to an increased overconfidence in retrospective judgments in comparison to prospective judgments.

Taken together, these studies provide remarkably consistent findings indicating that searching the internet for information seems to bias metacognitive judgments towards overestimating one’s own knowledge significantly in comparison to conditions when no internet is available.

It is helpful to conceptualise the tasks in these studies as complex information problem solving processes with the internet (i.e., IPS-I processes; Brand-Gruwel et al., 2009), in order to hypothesise about potential explanations for these results. IPS-I processes can be subdivided into distinct phases: (a) definition of the information problem, (b) searching information, (c) scanning information, (d) processing information, and (e) organising and presenting information. When engaging in such a complex task, a cognitive bias could potentially occur at different or even multiple points. For example, an overestimation bias might be enhanced because of a reliance on heuristic cues such as the fluency of the internet search or the ease of information retrieval that might be relevant during phase “(b) searching information” (Fischer et al., 2015; Pieschl, 2021; Dunn et al., 2021).

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

With the present study we explore a complementary perspective, namely the hypothesis that such an IPS-I overestimation bias might be based in phase "(a) definition of the information problem" when complex and ill-structured explanatory knowledge tasks are considered.

Standards in Explanatory Knowledge

Most often cognitive biases in metacognitive self-assessment are attributed to insufficient processing capabilities (Haselton, Nettle & Murray, 2015) or information-based and/or experience-based cues (Koriat et al., 2008; Koriat & Levy-Sadot, 1999). Extending this perspective, we will explore whether a mismatch between external task demands and subjective idiosyncratic task definitions could also contribute to such metacognitive overestimation biases. In other words, people may base their metacognitive judgments on insufficient criteria, i.e. they may use different standards than intended by external task demands or evaluations.

Especially in complex ill-structured tasks like providing explanations for natural phenomena, people's metacognitive judgments and idiosyncratic task definitions are usually based on their own explanatory understanding and personal mental models (Keil, 2006; Lombrozo, 2006). As explanations may be skeletal, incomplete, full of gaps in nature and are of the built "on the fly", a person may not recognize the state of their own knowledge. This is because explanations in themselves have no clearly defined end state or "objective truth" (Rozenblit & Keil, 2002). While many factual questions may have one clearly defined correct answer, explanations can be correct (or incorrect) to varying degrees and at different levels: for example, one could provide a (relatively) correct general explanation of the human circulatory system to e.g. first graders, while a student of medicine might write a detailed PhD thesis about just a small fraction of the circulatory system. Therefore, metacognitive judgments about one's own explanatory knowledge cannot be compared to an objective "ground truth" but rather reflect participants' – and researchers' – own task definitions.

The COPES model by Winne and Hadwin (1998) characterises the link between metacognitive judgments and knowledge as follows: Individually set standards are described as an optimal end-state or success state that determines behaviour in a goal-directed manner. Any product of such an intended effort is compared to these internal standards by the act of (metacognitive) monitoring. Regarding metacognitive judgments about explanatory knowledge tasks, it is not clear, by default, which frame of reference (Pieschl, 2009) a judgment is made against or which idiosyncratic internal standard a specific person has set. Additionally, as standards are internal in nature they do not necessarily match external evaluation criteria (Winne, 2021).

The relation of effort regulation and metacognition can, in this way, be described by discrepancy reduction models, where differences between a current state and an internal stopping criterion then guide goal-directed behaviour (Metcalfe, 2009). As monitoring and control of cognitive processes and behaviour depend on metacognition and metacognitive judgments (Bjork, Dunlosky, & Kornell, 2013; Dunlosky & Metcalfe, 2008; Koriat, 2018; Nelson & Narens, 1994), any effort regulation or effort control would also be affected by such internal standards. Applied to an explanatory knowledge task, the question becomes, when are people satisfied with an explanation or when do they think an explanation is sufficient? For example, when an internal criterion or standard is, in this regard, less strict than an external objective or evaluation criterion, an explanation may appear insufficient from an external perspective but might be judged appropriate from a subjective perspective. This is especially interesting in relation to IPS-I tasks because the internet opens up the possibility for additional knowledge related searches that could expand the current knowledge base.

Thus, we can draw the following conclusions for complex ill-structured explanatory knowledge tasks: Responders' internal task definitions – and therefore also their internal standards – may be idiosyncratic and might differ substantially from external task demands assumed by experimenters. Thus, metacognitive judgments may appear to indicate substantial overestimation. This hypothesis can be tested by experimentally aligning participants' internal task demands and standards with evaluators' external evaluation criteria.

The Present Study

To test our central hypothesis, we extended the research paradigm of Pieschl (2021). As in previous research, we included internet and no-internet conditions for the information problem solving tasks of answering explanatory knowledge questions. In these conditions, as in previous research, participants' metacognitive judgments were based on their idiosyncratic internal standards, that is, they had to use their own frame of reference for metacognitive judgments, and they were naïve to how their explanatory answers would be analysed. Extending previous research, we included additional conditions that aimed at aligning participants' internal standards with external evaluation criteria. In these conditions, we provided participants with an external rubric-schema clarifying how answers would be scored, that is, we provided an external frame of reference for their metacognitive judgments. The same rubric-schema was used later on by independent external evaluators to score participants' answers.

We tested the following hypotheses. Hypothesis 1 (H1): When participants use the internet to answer explanatory knowledge questions, their metacognitive confidence judgments indicate significantly more overestimation of their performance than without internet use. We tested H1 only in conditions without any rubric-schema (replication of Pieschl, 2021). Hypothesis 2 (H2): When participants are provided with external standards (i.e., the rubric-schema), their metacognitive judgments are more accurate than when not providing such standards. In addition to this, this, the research design allowed for the exploration of an interaction effect between the internet and standards conditions.

Methods

Participants

In this paper we include data from the (partially) overlapping parts of two studies. Both online convenience samples were collected via university mailing lists, social media groups and by word-of-mouth recommendations. All participants were informed about the procedure of the respective study and gave informant consent beforehand. Both studies were approved by the responsible ethics committee of the Technical University of Darmstadt.

Sample 1 consists of N = 84 participants and Sample 2 consists of N = 82 participants. Sample 1 participants were compensated by a voluntary participation in a ruffle of vouchers of 50€ and 15€ for every fifth person. Sample 2 participants were compensated with 15€ each. Sample 2 participants were rewarded more compensation because of an additional requirement to record their computer screens during the procedure of the experiment.

Taken together (Sample 1 + Sample 2) we report data from N = 166 participants. The total sample consists of 73% female, 25% male, 1% diverse, and 1% did not specify their gender. The age ranged from under 21 to over 65 with a median age of 26-30 years. Participants were mostly people "in school or vocational training" (42%) and "students" (28%). The majority indicated to use the internet at least "several times per day" (43%) or "daily" (28%).

Design and Procedure

In this paper we report a 2x2 within-subjects manipulation with the factors Internet (Internet vs. No-Internet) and Standards (Standards vs. No-Standards). The procedure of the experiment is displayed in Figure 1. First, participants were randomly assigned to answer two explanatory knowledge questions, one with and one without the internet without any additional instruction (No-Standards condition). This served as the direct replication of the main hypothesis from Pieschl (2021). Then, participants were instructed in using the rubric-schema (Standards conditions) and again answered two explanatory knowledge questions, one with and one without the internet in random order[1]. Prior as well as after participants answered each explanatory knowledge question they provided metacognitive as well as other judgments[2]. The design, hypotheses as well as the exclusion criteria were preregistered.

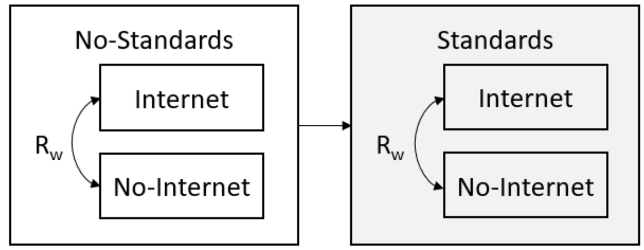

Figure 1: Experimental procedure (from left to right). In total, participants had to answer four explanatory knowledge questions. First, participants had to answer two questions without any additional instructions (No-Standards), one with and one without the internet in random order. After that, participants were additionally instructed with a rubric-schema (Standards) and again answered two questions in random order. R_w = within-subject randomization.

Explanatory Knowledge Questions

Participants had to give written and open-ended answers to a total of four out of six explanatory knowledge questions that were originally introduced by Fisher et al. (2015). The questions were picked at random during the experiment and are depicted here from the hardest to the easiest question (based on the item difficulties): "How do scientists know that the universe is expanding?", "Why does Swiss cheese have holes?", "Why are cloudy nights warmer?", "How do tornadoes form?", "How do scientists determine the dates of fossils?", and "How does the heart pump blood?".

Internet and Standard Conditions

In both, Internet and No-Internet conditions, participants had 15 minutes in total to write an answer to an explanatory knowledge question. In the Internet condition, participants were instructed to freely browse the internet to access any resource to answer the question. They were instructed to not copy and paste any information from websites but had to use their own words. In the No-Internet conditions, participants were instructed not to use any additional help or resource but to solely rely on their own knowledge.

For the Standard conditions, participants were explicitly instructed to base their metacognitive judgments on a newly developed rubric-schema. In the No-Standards conditions participants had to rely on their own frame of reference. The same rubric-schema was afterwards also used by the independent raters to judge the performance of the participants.

For the development of the rubric-schema we referred to three lines of previous research, namely the concepts of "explanatory depth" by Rozenblit and Keil (2002) and Keil (2006), "conceptual complexity" by Brown (2005) and Brown and Wilson (2011), and "knowledge integration" by

[1]In the study of Sample 1, we also tested a between-subjects manipulation of Standards. We only report data from the within-subjects manipulation here but obtained comparable results with the between-subjects manipulation of Standards.

[2]Not reported here.

Liu et al. (2008) and Lee, Liu and Linn (2011). Our rubric-schema had 7 categories, from 1 as the category with lowest complexity ("The description contains isolated concepts or elements, but no interactions between the elements are listed. No relationships or dependencies between the elements are shown. Circumstances or initial conditions under which the relationships between the elements come into play are completely absent") to 7 as the category with highest complexity ("The description contains all the important concepts or elements and shows how the mechanisms of action between the elements are mutually interdependent. It is described how elements depend on or interact with each other and thus together create an integrated overall picture of the phenomenon. Circumstances or initial conditions under which the relationships between the elements come about are given."). We added the category 0 ("no or only irrelevant information"), had additional descriptives for 3 and 5 and had empty categories at the steps of 2,4 and 6 for answers falling in between two descriptive levels.

Measures

Performance For each sample, two independent raters who were blind to the experimental conditions rated the answers of participants to the explanatory knowledge questions using the developed rubric-schema. As an initial training step, the raters evaluated 5% of the written answers together and afterwards independently evaluated all written answers. The average of both ratings determined participants' Performance score regarding a particular question. Inter-rater agreement was high, with a Krippendorff's Alpha of .930 for Sample 1 and .926 for Sample 2 (with a perfect match at 1 and an acceptable alpha above .80; Krippendorff, 2004). One of the raters was present in both samples.

Self-reported judgments Before and after answering each explanatory knowledge question, participants had to give predictive ("How confident are you that you will give a good answer to this question?") and postdictive ("How confident are you that you gave a good answer to this question?") Metacognitive Confidence Judgments (see also Pieschl, 2021) for which we had pre-registered hypotheses. We used a visual analogue scale ranging from 1 = *not at all confident* to 7 = *very confident* to record their answers. Participants could also indicate *not specified*.

Bias-score To quantify the accuracy of participants' Metacognitive Confidence Judgments, an absolute Bias-score was calculated by subtracting each Performance score from the corresponding Metacognitive Confidence Judgment (Nelson & Narens, 1984; Schraw, 2009). A positive Bias-score indicates overestimation, a negative Bias-score underestimation, and Bias-scores around zero indicates accurate calibration.

Screen recordings In the case of Sample 2 we additionally collected screen recordings of participants. This gave us the possibility to check whether or not participants complied with the experimental instructions (Internet and No-Internet)[3].

Power Considerations and Analysis

To replicate the main Hypothesis by Pieschl (2021) we calculated an estimated sample size between 31 participants (with $\alpha = .05$, power = .9 and $d = 0.55$ for the original effect size) and 97 participants (with $\alpha = .05$, power = .9 and $d = 0.3$ for a more conservative estimate). We went with the more conservative estimate as a goal because of unknown effect sizes for our second hypothesis.

We preregistered to compute ANOVAs (or equivalent non-parametric analyses) for the planned comparisons to answer our hypothesis. However, we acknowledge that we deviated from the preregistered analysis plan for testing the first hypothesis (H1) because a *F*-test for equal variances revealed, that the conditions had a variance ratio unequal to 1, with $F(165, 165) = 0.54, p < .001$ for the predictive and $F(165, 165) = 0.59, p < .001$ for the postdictive phase, respectively. But since the Bias-scores were nearly normally distributed, we reasoned that a paired *t*-test would be appropriate to compare the Internet and No-Internet conditions under these circumstances.

To control for any effect of sample (e.g. systematic difference in Bias-score in Sample 1 and Sample 2), we also included Sample as a factor in our analyses.

Results

To test the first hypothesis (H1), we compared Internet and No-Internet conditions via paired *t*-tests (only No-Standards conditions), which was a direct replication of the central hypothesis from Pieschl (2021). In the predictive phase, there was a significant difference between Internet ($M = 0.60$, $SD = 2.14$) and No-Internet ($M = 0.25$, $SD = 1.58$) conditions, with $t(165) = 1.88, p = .031, d = 0.19$. In the postdictive phase, there was also a significant difference between Internet ($M = 0.55, SD = 2.10$) and No-Internet ($M = 0.25, SD = 1.61$) conditions, with $t(165) = 1.82, p = .035, d = 0.16$. These results are visualised in Figure 2 under "No-Standards".

Regarding the second hypothesis (H2), we computed a 2x2x2 ANOVA with the factors Internet, Standards, and Sample and found a significant main effect of Standards on the Bias-score in the predictive as well as in the postdictive phase (see Table 1). The manipulation of participants' standards led to an overall decrease in the Bias-score and to more accurate metacognitive judgments, especially regarding the postdictive judgments in the Internet condition (see Figure 2). In this analysis across all conditions, the main effect for Internet was not significant in the predictive phase; however, there was a significant main effect for Internet in the postdictive phase. No interaction effect was found

[3]The screen recordings were also used to code for the distinct phases of the IPS-I process. However, the process analysis is not part of this paper.

between Internet and Standards. Additionally, there were no significant differences between Samples or significant interaction effects including the factor Sample.

Discussion

This study replicates previous findings of participants overestimating their ability in explanatory knowledge tasks involving internet use compared to no-internet use (Pieschl, 2021). We showed that when participants had to rely on their own frame of reference, they were biased in the direction of overconfidence when answering explanatory knowledge questions. These results are in accordance with previous studies' findings, reporting heightened metacognitive confidence in knowledge tasks involving the internet (Pieschl, 2021; Fischer et al., 2015). It is therefore important to determine the cognitive and behavioural processes that might influence this phenomenon. If a systematic overestimation in tasks involving knowledge and the internet occurs, then this could have, for example, direct consequences for formal and informal education, where using the internet becomes more and more ubiquitous (Pieschl, 2021). Specifically, with increased confidence, participants could be more likely to abandon tasks too early and could be prone to accepting inadequate answers as sufficient.

The present study further showed that a divergence of applied criteria or standards for task completion (Pieschl, 2009) could be one potential variable causally contributing to such overconfidence. A closer alignment of the reference frames of the external evaluators and participants seems to have reduced the metacognitive overestimation bias. The introduction of external standards substantially changed the direction and magnitude of metacognitive Bias-scores. With such alignment, participants' metacognitive confidence judgments were accurately calibrated to their performance in the internet condition.

Without explicit communication of standards, participants will likely default to their own personal definitions of what a "good" explanation is. These internal standards can differ widely between people (Ryan, 1984), as such standards might be very well-suited for their particular personal context.

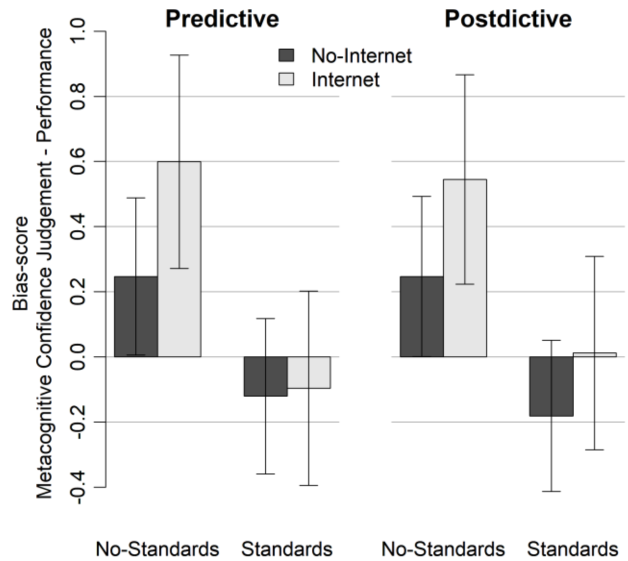

Figure 2: Predictive and postdictive Bias-scores by experimental conditions. Error bars represent 95% confidence intervals. The Bias-score was calculated by subtracting Performance from the corresponding Metacognitive Confidence Judgment. A positive value indicates overconfidence, a negative value underconfidence, a value around 0 indicates accurate judgments. The No-Standards conditions constitute a replication of Pieschl (2021).

However, they may not match external evaluation criteria (Winne, 2021). The reference to adequate standards is, therefore, especially important for complex ill-structured tasks such as explanatory knowledge information problem solving tasks. Thus, our results highlight that the metacognitive overestimation bias detected in previous studies with explanatory knowledge tasks could also be attributed to the very first phase of the IPS-I process, namely in the definition of an information problem.

In the overall ANOVA controlling for the standard conditions, the main effect of the Internet was not significant

Table 1. 2x2x2 ANOVAs of the predictive and postdictive Bias-score.

	Predictive Bias-score				*Postdictive Bias-score*			
Predictor	df_n, df_d	*F*	*p*	η_p^2	df_n, df_d	*F*	*p*	η_p^2
Within-subjects effect								
Internet	1, 164	1.92	.168	0.01	1, 164	4.40	.037	0.03
Standards	1, 164	18.69	< .001	0.10	1, 164	16.22	< .001	0.09
Sample	1, 164	3.08	.081	0.02	1, 164	1.93	.166	0.01
Internet x Standards	1, 164	2.00	.159	0.01	1, 164	0.26	.613	0.00
Sample x Internet	1, 164	0.19	.663	0.00	1, 164	0.18	.672	0.00
Sample x Standards	1, 164	2.06	.153	0.01	1, 164	3.70	.056	0.02
Sample x Internet x Standards	1, 164	0.01	.921	0.00	1, 164	0.10	.757	0.00

Note: The factor Sample was included to control for any systematic difference in the two samples. However, the results of the factors Internet and Standards did not change when the factor Sample was removed from the analysis.

in the predictive phase. This shows once again the effect of standards on the IPS-I task: With explicit standards participants in our sample could fairly well predict their performance – with and without the internet. However, the tendency of participants to descriptively underestimate their performance in the No-Internet and Standard condition in the postdictive phase may have been the reason why the main effect of Internet was significant in the postdictive phase in the overall ANOVA.

We speculate that one explanation why participants descriptively underestimated their performance in the Standards condition of the No-Internet and not in the Internet conditions is that participants might have been overly critical about their comparatively worse performance without internet. Our pool consisted of participants who used the internet frequently and they therefore might already be well habituated to use the internet in knowledge related tasks (Sparrow et al., 2011; Ward, 2013). Thus, it could be assumed that participants would better be able to accurately judge their performance when they had access to the internet – at least in their postdictive metacognitive judgments. In contrast, having to rely solely on their own knowledge may be an atypical situation for internet users, especially in such close proximity to internet access and without the possibility to review or validate their own answers. However, as the underestimation was not statistically significantly different from zero, further studies are needed to explore this point further.

Limitations

Our sample consisted mostly of young participants that used the internet very frequently. This limits the generalisability of the results as the heterogeneity of the sample was limited. Also, this study was conducted in an online format and, thus, we could not ensure full control over situational aspects that might have influenced the experimental outcome. For example, we do not know if participants used other external knowledge besides the internet or used other devices to access the internet entirely. We wanted to minimise other factors in the second sample of this study. Therefore, we additionally instructed participants to record their computer screen during the experiment. This gave us the opportunity to screen the video recordings of the experimental procedure for any behaviour that was not intended by our instruction. Crucially, in the second sample we identified six participants who did not follow the instruction of using or not using the internet, as well as seven participants that used the internet prior to the self-judgments in the predictive phase. Based on the screen-recordings we were able to exclude these participants from Sample 2 and hence were better able to apply our exclusion criteria. However, this was not possible for Sample 1 without the screen recordings. Nonetheless, we estimate that this did not substantially influence our results as our ANOVA revealed no significant effects of Sample or any significant interactions involving Sample for the Bias-scores (see Table 1).

The dependent variable of the Bias-score was calculated from a relation between performance and confidence. Since it is known from other contexts that rubrics can increase performance (cf. Howell, 2014), we also exploratively checked whether this was the case here. However, there were no significant differences in Performance between Standards and No-Standards conditions. The results reported here are therefore actually due to changes in confidence ratings as a result of the developed rubric-schema.

In this experiment, participants were not free to choose how much time they wanted to spend answering an explanatory knowledge question because the time was limited to 15 minutes per question. This might have induced different time pressures in the Internet compared to the No-Internet condition because of the additional time needed for the internet search process.

Conclusion

With explicit communication of evaluative standards, participants of this experiment were fairly accurately calibrated in their metacognitive judgments about their performance in the explanatory knowledge task. On a theoretical level, this finding contributes to explanations about the roots of the previously observed metacognitive overestimation bias in IPS-I tasks, suggesting that people may already derive idiosyncratic – and potentially biased – definitions of the information problems at the start of the search process. Of course, this finding does not imply that later phases of the IPS-I process might not play a part in a potential metacognitive bias as well. For example, experiential cues such as fluency of internet searches could also be contributing to the overestimation. On an applied level, providing people with external standards, as done in this experiment, may be a feasible buffer against metacognitive overestimation biases in IPS-I tasks. Teachers may clearly communicate their requirements whenever they task students with IPS-I in order to optimise students' self-assessments and thereby also their metacognitive regulation.

References

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual review of psychology*, *64*, 417-444.

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Comput. Educ., 53*, 1207-1217.

Brown, N. J. S. (2005). *The multidimensional measure of conceptual complexity*. Berkeley: University of California.

Brown, N. J., & Wilson, M. (2011). A model of cognition: The missing cornerstone of assessment. *Educational Psychology Review, 23*(2), 221.

Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. Sage Publications.

Dunn, T.L., Gaspar, C., McLean, D., Koehler, D.J., & Risko, E.F. (2021). Distributed metacognition: Increased bias and deficits in metacognitive sensitivity when retrieving information from the internet. *Technology, Mind, and Behavior*.

Firth, J., Torous, J., Stubbs, B., Firth, J. A., Steiner, G. Z., Smith, L., Alvarez-Jimenez, M., Glleson, J., Vancampfort, D., Armitage, C. J., & Sarris, J. (2019). The “online brain”: how the Internet may be changing our cognition. *World Psychiatry, 18*(2), 119-129.

Fisher, M., Goddu, M.K., & Keil, F.C. (2015). Searching for explanations: How the Internet inflates estimates of internal knowledge. *Journal of experimental psychology. General, 144 3*, 674-87 .

Greene, J.A., Cartiff, B.M., & Duke, R.F. (2018). A Meta-Analytic Review of the Relationship Between Epistemic Cognition and Academic Achievement. *Journal of Educational Psychology, 110*, 1084–1111.

Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The evolution of cognitive bias. *The handbook of evolutionary psychology*.

Howell, R. J. (2014). Grading rubrics: hoopla or help?. *Innovations in Education and Teaching International*, *51*(4), 400-410.

Keil, F. C. (2006). Explanation and understanding. *Annu. Rev. Psychol.*, *57*, 227-254.

Koriat, A. (2018). When Reality Is Out of Focus: Can People Tell Whether Their Beliefs and Judgments Are Correct or Wrong? Journal of Experimental Psychology: General, 147, 613–631.

Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Informationbased and experience-based monitoring of one’s own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology*. New York: Guilford Publications.

Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. *A handbook of memory and metamemory*.

Krippendorff, K. (2004). Content analysis: An introduction to its methodology. Thousand Oaks, California: Sage.

Lee, H. S., Liu, O. L., & Linn, M. C. (2011). Validating measurement of knowledge integration in science using multiple-choice and explanation items. *Applied Measurement in Education, 24*(2), 115-136.

Liu, O. L., Lee, H. S., Hofstetter, C., & Linn, M. C. (2008). Assessing knowledge integration in science: Construct, measures, and evidence. *Educational Assessment, 13*(1), 33-55.

Loh, K.K., & Kanai, R. (2016). How Has the Internet Reshaped Human Cognition? *The Neuroscientist, 22*, 506-520.

Lombrozo, T., Sloman, S., Strevens, M., Trout, J. D., & Weisberg, D. S. (2008). Understanding Why: The cognitive science of explanation. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 30, No. 30).

Marsh, E. J., & Rajaram, S. (2019). The digital expansion of the mind: Implications of internet usage for memory and cognition. *Journal of Applied Research in Memory and Cognition*, *8*, 1-14.

Metcalfe, J. (2009). Metacognitive Judgments and Control of Study. *Current Directions in Psychological Science, 18*, 159 - 163.

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition. *Metacognition: Knowing about knowing*, *13*, 1-25.

Pieschl, S. (2009). Metacognitive calibration—an extended conceptualization and potential applications. *Metacognition and Learning, 4*, 3-31.

Pieschl, S. (2021). Will using the Internet to answer knowledge questions increase users’ overestimation of their own ability or performance? *Media Psychology, 24*, 109 - 135.

Rozenblit, L., & Keil, F.C. (2002). The misunderstood limits of folk science: an illusion of explanatory depth. *Cognitive science, 26 5*, 521-562.

Ryan, M.P. (1984). Monitoring Text Comprehension: Individual Differences in Epistemological Standards. *Journal of Educational Psychology, 76*, 248-258.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4*, 33-45.

Sparrow, B., Liu, J., & Wegner, D.M. (2011). Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. *Science, 333*, 776 - 778.

Ward, A. F. (2013). *One with the Cloud: Why people mistake the Internet's knowledge for their own*. Harvard University.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice*. Hillsdale, NJ: Lawrence Erlbaum.

Winne, P.H. (2021). Cognition, Metacognition, and Self-Regulated Learning. *Oxford Research Encyclopedia of Education*.