



# Putting post-decision wagering to the test: a measure of self-perceived knowledge in basic sciences?

Marjolein Versteeg<sup>1,2</sup> · Paul Steendijk<sup>1,2</sup>

Published online: 5 February 2019  
© The Author(s) 2019

## Abstract

**Introduction** Students learn more effectively when they know what they do not know. Gaining insight into students' metacognitive awareness is needed as misalignment between actual and self-perceived knowledge impedes their learning process. The optimal method of measuring self-perceived knowledge is still under debate. In this study, we evaluate the use of psychology-derived post-decision wagering for mapping students self-perceived knowledge.

**Methods** Students ( $n=71$ ) performed a pre-test on medical physiology, followed by a teacher-moderated discussion and a post-test with isomorph questions. Half of the students rated their self-perceived knowledge on each question using post-decision wagering, i.e. betting 1–5 points on the correctness of their answer, whereas the other half used a 5-point Likert scale to rate their confidence.

**Results** Self-perceived knowledge scores were higher for post-decision wagering (pre:  $3.75 \pm 0.14$ , post:  $4.60 \pm 0.07$ ) compared with Likert scales (pre:  $3.13 \pm 0.08$ , post:  $3.92 \pm 0.08$ ) despite similar actual knowledge scores. Furthermore, Likert ratings showed a near-normal distribution, whereas wagers were placed preferentially using the outer ends of the scale. Correlations between mean actual and self-perceived knowledge scores were low in both groups. On average, 8.5% of responses were classified as misconceptions, defined as highly confident incorrect answers.

**Discussion** Despite the presumed reliability of post-decision wagering, our findings suggest that we should adhere to the use of Likert scales as a balanced measure for self-perceived knowledge in medical education. Moreover, the prevalence of misconceptions did not alter after instruction, indicating a need for instructional designs that enhance students' conceptual understanding in basic sciences.

**Keywords** Medical education · Instructional methods · Quantitative Research Methods · Metacognition

## What this paper adds

The importance of 'knowing what you do not know' is well-acknowledged in medical education. However, the optimal method of measuring self-perceived knowledge is still under debate. In this study, a psychology-derived tool to measure self-perceived knowledge called post-decision wagering was evaluated and considered less effective for educational settings than traditional Likert scales. We recommend including Likert scales in multiple-choice formats to establish confidence-weighted practice tests. Such tests

may inform both students and educators on the student's level of understanding, making it a more powerful tool for improving teaching and learning.

## Introduction

Learning basic sciences in medicine is recognized as an important, but challenging undertaking. It requires students to achieve understanding of complex learning material by analyzing, conceptualizing, and integrating knowledge. Strikingly, students are often unaware of what they do and do not know, which can be detrimental to learning [1, 2]. We cannot expect students to perform as effective learners if they are unaware of their own knowledge deficiencies.

Being aware of knowledge deficiencies is considered part of one's metacognition. Metacognition, put simply, refers to thinking about one's own thinking [3]. The ability to evaluate one's knowledge (i.e. self-perceived knowledge) after

✉ Marjolein Versteeg  
m.versteeg@lumc.nl

<sup>1</sup> Center for Innovation in Medical Education, Leiden University Medical Center, Leiden, The Netherlands

<sup>2</sup> Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands



performing a task or test (i.e. actual knowledge) is a sub-component of metacognition, which we refer to as *metacognitive evaluation*: knowing how well you did after finishing a task [3–5]. Metacognitive evaluation is considered a critical component for learning as it informs both students and educators on students' potential knowledge deficiencies [6, 7]. Moreover, some researchers have showed that including metacognitive evaluation through confidence marking during multiple-choice practice tests increases the predictive value of testing [8–10].

However, studies assessing metacognitive evaluation in medical education mainly focus on clinical knowledge [11–18], while the domain of basic sciences remains largely unexplored. Particularly within this latter domain, misunderstandings of physiological scientific concepts should be detected and corrected, since these concepts form a foundation for accurate clinical reasoning [19, 20]. Regarding basic science knowledge, Dawson et al. illustrate that according to faculty teachers physiological topics are experienced by medical students as being among the most difficult [21]. This finding is supported by research that has shown a high prevalence of misunderstandings among students in medical physiology education [22–25]. According to conceptual change theory, these so-called *misconceptions* are different from a mere lack of knowledge; misconceptions are robust to change and are therefore difficult to alleviate by traditional teaching methods [26]. Misconceptions are generally indicated by incorrect answers that are given by students with high confidence, indicating that students are unaware of their incorrect knowledge [27]. The role of awareness in establishing conceptual understanding illustrates that conceptual change theory and metacognitive theory are strongly intertwined in practice. Recognizing and evaluating existing conceptions, and deciding whether to reconstruct or review one's understanding are all metacognitive processes [28]. Conceptual change theory emphasizes the key role of awareness in accomplishing the shift from a misconception to the scientifically correct conception [26, 29–31]. Awareness can be created by explicitly assessing students' metacognitive evaluation in the classroom, as it forces both learners and educators to think about their understanding of the subject matter.

Research on science learning in higher education nicely illustrates how metacognitive evaluation can be measured in classroom settings using a multi-tier approach [32–34]. Multi-tier assessments consist of multiple-choice exercises with a complementary Likert scale question: 'How confident are you that your answer to this question is correct?'. This is an example of measuring metacognitive evaluation; assessing students' actual knowledge through test scores and their self-perceived knowledge through Likert scales. The use of Likert scales as a self-report tool for self-perceived knowledge is very common in educational research,

with variations in scales ranging from 3 to an increasing number of integers [11–15, 32–35]. However, there is some controversy on the objectiveness of Likert scales for measuring self-perceived knowledge. As noted in a review by Koch and Preuschoff: 'Studying consciousness is difficult because asking subjects to report their awareness of a stimulus perturbs that awareness' [36]. In psychology research, an alternative method to more implicitly measure awareness of visual stimuli was introduced, called post-decision wagering (PDW) [37]. This method is suggested to determine awareness more objectively by indirectly assessing individuals' self-perceived knowledge by betting points (or money) on their decisions. If an individual chooses the correct answer points are gained, whereas with a bet on an incorrect answer points are lost. Additionally, researchers suggest that PDW enhances individuals' motivation to reveal their confidence in their answer compared with self-reports on numerical confidence scales [38, 39]. Contrastingly, various studies show a substantial influence of loss aversion in PDW resulting in a decreased accuracy towards measuring self-perceived knowledge. Therefore, participants are less prone to indicate a high confidence level through wagering [40, 41]. Moreover, wagers are also suggested to depend on wager size indicating that PDW is still a subjective tool to measure awareness [42].

The use of PDW as a tool to measure self-perceived knowledge has not yet been reported in medical education. In this study, our primary objective was to evaluate psychology-derived PDW as a measure of self-perceived knowledge in educational contexts. We compared PDW with Likert scales and hypothesized that confidence wagers rather than ratings would be more aligned with students' actual knowledge due to their supposed more objective nature. To further determine the practical usefulness of the instruments in quantitating self-perceived knowledge, we examined the distribution of responses along the scales to compare effective resolution and discriminative power. As a secondary objective, we investigated the prevalence of misconceptions among students to gain insight into their conceptual understanding of the subject matter. In the present study, we focused on basic sciences particularly given the high prevalence of misconceptions [21–25, 43]. Misconceptions may not only affect students' conceptual understanding, but also have significant impact on their clinical reasoning skills [23, 44, 45], and are therefore considered an important subject of investigation in medical education.

## Methods

### Participants

A cohort of first-year bachelor Biomedical Sciences students ( $n = 71$ ) from Leiden University Medical Center participated in this study. In the Netherlands, similar physiology courses are part of both the biomedical and medical bachelor curriculum.

### Setting

For this study, the online platform *Socrative* was used to develop multiple-choice knowledge tests. The knowledge tests were implemented in a compulsory 2-hour supervised seminar on cardiovascular physiology that was taught in five small working groups (13–15 students/group) by the same teacher. Allocation of the students to the working groups was arbitrary except for the aim to have a similar female/male ratio in all groups. The topics were introduced in a preceding plenary lecture, and students were provided with a handout during the seminar which contained some basic figures and diagrams. Although the knowledge tests were a mandatory part of the seminar, students could voluntarily decide if their anonymous answers could be used for educational research purposes by giving informed consent. No incentives were offered for participation and test results did not affect the course grade. Students could withdraw their permission at any time. This experiment was part of a larger study also investigating the influence of peer discussion on confidence of which data are not reported in this paper. This study was approved by the Educational Research Review Board (ERRB) of Leiden University Medical Center: ERRB reference number: OEC/ERRB/20180612/3.

### Procedure

The knowledge tests were taken individually under exam conditions, at the beginning (pre) and at the end (post) of the seminar. Both tests consisted of 10 multiple-choice questions and examined the same knowledge base. The post-test used slightly different phrasing of the questions which can be considered isomorphic or so-called near-transfer questions. All students were instructed to answer each question individually and to provide a confidence rating (Likert scale: 3 out of 5 groups,  $n = 42$ ) or wager (PDW: 2 out of 5 groups,  $n = 29$ ) immediately after each question. To compute the actual knowledge, students received one point for every correct answer of the multiple-choice tests yielding a maximum score of 10 points per test. During the body of the actual seminar the questions from the pre-test and related topics were explained and discussed.

Questions were framed according to the revised version of Bloom's taxonomy of cognitive domain and all categorized as comprehension-type questions. To ensure validity of the knowledge test, the physiology questions were designed by an expert physiologist (PS) and derived from a database of previously used exam questions. PS is an educational professor at Leiden University Medical Center and actively involved in shaping the biomedical and medical curriculum regarding medical physiology education, and familiar with the learning goals of the human biology course. Consequently, the questions were designed in such a way that they would meet the overall course objectives regarding physiological knowledge.

### Instruments

For confidence rating in the Likert scale group (LS group), students used a 5-point Likert scale rating instrument: (1) Completely unsure (just guessing), (2) Rather unsure, (3) Neutral (50/50), (4) Rather sure, (5) Very sure (almost 100%). Students received additional instruction on the neutral item (3) as the lecturer indicated that this rating should be given when a student doubted between two remaining answer choices considered equally likely. In the PDW group, students were instructed to place their bets on the correct answer. In order to compare results with the LS group, wagers ranged from 1 to 5 points. Students could bet 1–5 points per question, which were gained if the answer was correct and lost if the answer was incorrect. Students received their total wagering scores after the seminar.

### Data analyses

Descriptive statistics are provided as means and standard errors of the mean, unless otherwise mentioned. Actual and self-perceived knowledge scores in PDW and LS groups were compared using independent samples t-tests. Com-

		Actual knowledge	
		Incorrect	Correct
Self-perceived knowledge	High	Misconception	Correct knowledge
	Low	Lack of knowledge	Lucky guess

Fig. 1 A decision matrix based on students' actual and self-perceived knowledge. Adapted from [27]



**Table 1** Mean scores for students' actual and self-perceived knowledge

Instrument	Actual knowledge (Score, max 10)	Self-perceived knowledge (Score, max 5)	No. of students
<i>Pre-test</i>			
Post-decision wagering	6.78 ± 0.27	3.75 ± 0.14	29
Likert scale	6.38 ± 0.27	3.13 ± 0.08	42
<i>Post-test</i>			
Post-decision wagering	8.31 ± 0.23	4.60 ± 0.07	29
Likert scale	7.68 ± 0.27	3.92 ± 0.08	41

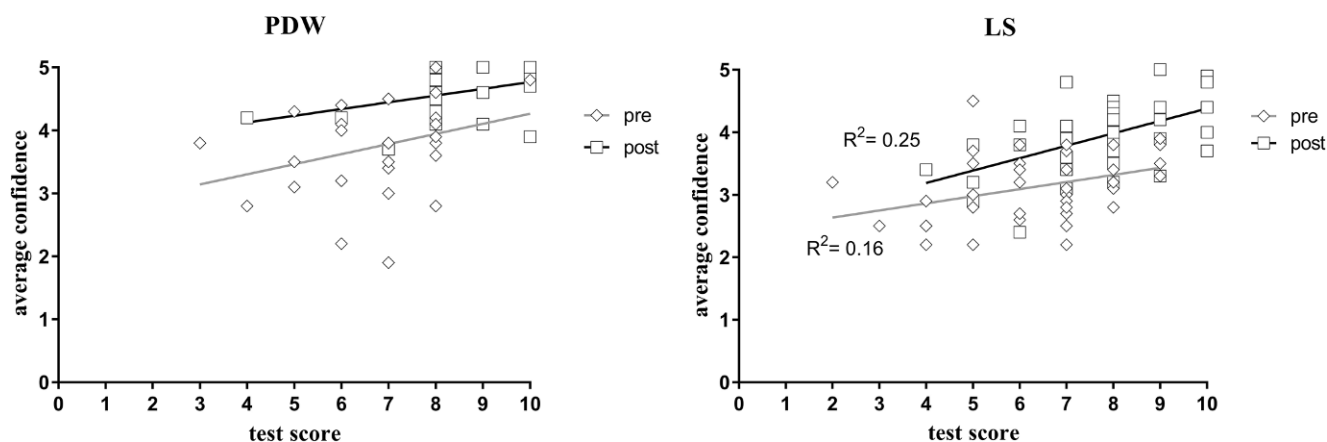
parison of score gains on the pre-post-test between groups was performed with a repeated measures ANOVA. To test and compare alignment between actual knowledge and self-perceived knowledge scores, correlation analysis was performed using Spearman's rank order correlation coefficient. Scores were averaged per student. A Fisher's  $r$ -to- $z$  transformation was performed to test for potential significant differences between correlations. The discriminative power of both methods was tested by performing Chi-square tests of independence. Hasan's decision matrix was used as a template to analyze the prevalence of misconceptions (Fig. 1; [27]). To create dichotomous outcomes, a threshold value of 3 was adopted and used as a cut-off point. Students were instructed to rate an answer with 3 when they were in doubt between two answer choices considered equally likely, meaning that they were still unsure about knowing the correct answer, i.e. having low confidence. Correct answers self-rated with a high confidence ( $>3$ ) were classified as 'correct knowledge' and when rated with low confidence ( $\leq 3$ ) as 'lucky guess'. Incorrect answers with a low confidence rating were considered 'lack of knowledge' and when paired with high confidence, the response indicated the presence of a misconception. IBM SPSS Statistics Version 23.0 (IBM Corp., Armonk, New York, USA) and GraphPad Prism Version 7.02 (GraphPad Software, La Jolla, California, USA) were used for all data analyses and visualizations.

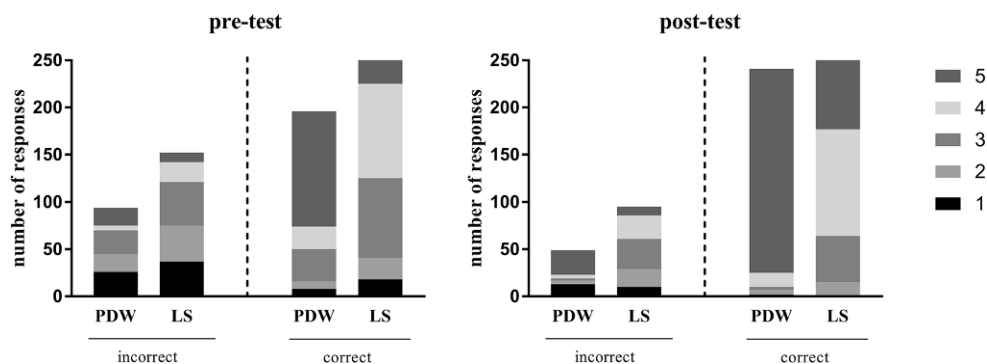
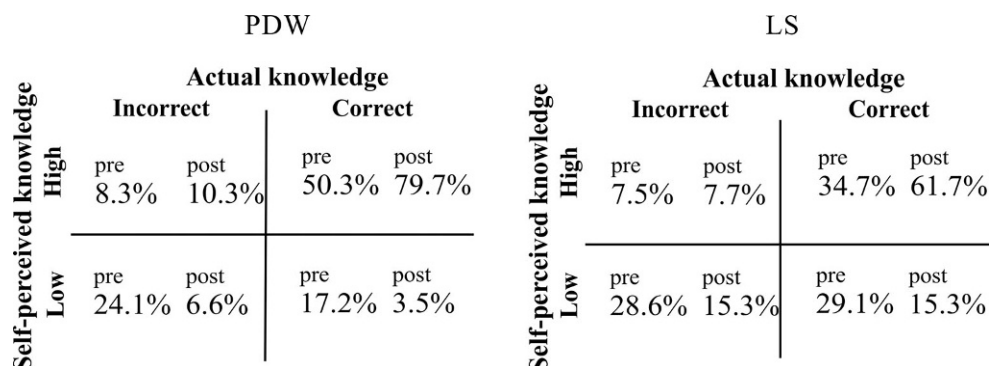
## Results

The average confidence scores reported in the PDW group on the pre-test and post-test were significantly higher compared with the scores in the LS group (both  $p < 0.001$ ) (Tab. 1). The gain in confidence scores in the PDW group ( $0.86 \pm 0.12$ ) was significantly higher than in the LS group ( $0.79 \pm 0.12$ ) ( $p < 0.001$ ,  $\eta^2 = 0.349$ ). The average test scores did not differ significantly between PDW and Likert scale on the pre-test ( $p = 0.338$ ) and post-test ( $p = 0.065$ ). In the PDW group, students showed an average increase of  $1.55 \pm 0.30$  points in their pre- to post-test scores, which was not significantly different from the gain of  $1.32 \pm 0.28$  points in the LS group ( $p = 0.094$ ).

## Alignment

There was a significant correlation between actual and self-perceived knowledge scores in the LS group on the pre-test ( $R^2 = 0.16$ ,  $p < 0.01$ ) and post-test ( $R^2 = 0.25$ ,  $p = 0.001$ ) (Fig. 2). In the PDW group correlations between actual knowledge and self-perceived knowledge are absent on both tests (pre:  $R^2 = 0.12$ ,  $p = 0.063$ , post:  $R^2 = 0.10$ ,  $p = 0.102$ ). To test for potential differences between the correlations in the PDW and LS group for the pre-test and post-test respectively, correlations were transformed into  $z$ -scores us-

**Fig. 2** Relationship between actual and self-perceived knowledge of students

**Fig. 3** Distribution of self-perceived knowledge scores**Fig. 4** Prevalence of misconceptions

ing Fisher's *r*-to-*z* transformation. The differences between correlations on the pre-tests ( $z = -0.23$ ,  $p = 0.817$ ) and post-tests ( $z = -0.90$ ,  $p = 0.370$ ) were not significant.

### Distribution

The distribution of confidence scores (1–5) for the Likert scale and PDW instruments is shown in Fig. 3. Comparing the distribution of 1–5 responses on the pre-test yielded a significant difference in distribution in self-perceived knowledge scores between PDW and Likert scale on questions with an incorrect response ( $p < 0.007$ ,  $\varphi = 0.240$ ). Of all questions answered incorrectly in the PDW group, 25.5% of responses were rated with level 5 confidence, compared with only 6% in the LS group. An even more pronounced difference between the Likert scale and PDW was found for correctly answered questions, yielding 43.8% level 5 confidence scores in the PDW group versus 16.3% in the LS group ( $p < 0.001$ ,  $\varphi = 0.483$ ). Discrepancies in the self-perceived knowledge distribution between Likert scales and PDW were also obtained for the post-test. The largest effect size was obtained for incorrect answers on the post-test ( $p < 0.001$ ,  $\varphi = 0.593$ ) with 53.1% level 5 responses in the PDW group and 9.7% in the LS group. For correct answers, 89.6% was associated with confidence level 5 in PDW versus 43.6% in the Likert scale ( $p < 0.001$ ,  $\varphi = 0.483$ ).

### Misconceptions

There was an increase in 'correct knowledge' in both groups after the plenary instruction and discussion (PDW: 50.3 to 79.7%; LS: 34.7 to 61.7%) (Fig. 4). The total number of 'lucky guesses' was reduced (PDW: 17.2 to 3.5%; LS: 29.1 to 15.3%) as was the amount of 'lack of knowledge' (PDW: 24.1 to 6.6%; LS: 28.6 to 15.3%). The prevalence of misconceptions was similarly present before and after instruction (PDW: 8.3 to 10.3%; LS: 7.5 to 7.7%).

### Discussion

The aim of this study was to evaluate psychology-derived post-decision wagering (PDW) as a measure of self-perceived knowledge by comparing it with the commonly used Likert scales. Despite similar levels of actual knowledge in both groups, students who used wagers indicated more confidence in their answers compared with students who used ratings. PDW confidence scores were also less evenly distributed compared with Likert scales and did not show a normal distribution, resulting in a less proportional use of the instrument. Contrary to our a priori hypothesis, PDW thus represents a less balanced measure of self-perceived knowledge than the traditional Likert scale.

Misalignment between actual and self-perceived knowledge was present in both groups, indicating that students





generally do not know what they know or do not know. Some students using PDW mentioned that they would easily go ‘all in’ and bet all their points even when they were not completely sure, and despite the rule that points were lost in case of an incorrect answer. Contrary to these findings, psychological research and behaviour economics report that individuals who wager generally show lower confidence compared with verbal confidence reports due to so-called risk aversion [41, 42]. The apparent absence of risk aversion in our students might be due to the use of imaginary ‘points’ as incentives. Interestingly, however, studies in laboratory settings have reported similar responses when using real versus imaginary incentives [37]. We suggest that the educational environment in which students are socially engaged might influence their response to risk aversion. The effects of classroom versus laboratory settings on the use of self-perceived knowledge measures await further investigation.

Our findings are in line with previous studies reporting discrepancies between students’ actual and self-perceived knowledge consistently across disciplines [1, 6, 15, 24, 46–49]. Kruger and Dunning have suggested that incompetence deprives an individual of the ability to recognize its shortcomings [2]. Based on this reasoning, our reported underestimation of performance is an unexpected finding. An explanation for this discrepancy may be that metacognitive evaluation of the specific materials in this study did not resemble the more global self-assessments mentioned in Kruger and Dunning’s studies. Such global judgements are, for example, self-competence (‘I feel I am able to ...’) and self-efficacy (‘how confident are you that you can ...’). These are prospective judgements and can therefore be considered part of one’s metacognitive knowledge, whereas metacognitive evaluation (as measured in the present study) is an on-line judgement which is made after one has performed a specific task [4, 5]. Students might be more cautious with estimating their actual scores after they have performed a task compared with a situation in which a global prospective judgement is requested. Furthermore, gender differences may have influenced our outcomes. Research has indicated that women tend to underestimate their academic capacities, as illustrated by the study by Ravesloot on progress testing in medical education [50]. Most of the students (70%) in this experiment were female, which might partly explain the relatively large percentage of correct low-confidence responses.

Overall, teacher instruction during the seminar reduced the percentage of lack of knowledge and lucky guesses and increased the number of correct responses, whereas the number of misconceptions was not greatly altered. Based on conceptual change theory and schema theory, we assume that the seminar discussion may not have benefitted the students with misconceptions, as the instructional de-

sign to achieve conceptual change should include more explicit activation of students’ prior knowledge. A learner’s understanding is facilitated by adding new information to an existing mental model or schema in the brain, which comprises relevant prior knowledge a learner has already obtained [51–53]. Cognitive neuroscientists have shown that new information can be added to schemas (i. e. stable neural network) faster when this information fits the prior knowledge [54, 55]. Thus, when a learner’s schema encompasses a misconception, it may be difficult for the instructor to teach the scientific conception using traditional teaching methods. By activating students’ prior knowledge, their current understanding of concepts and potential misconceptions become detectable for instructors [56, 57]. Future studies may focus on the use of conceptual change instructions, including activation of prior knowledge, to alleviate the number of misconceptions and enhance students’ conceptual understanding.

Our study has several limitations. We focused on the comparison of different measurement instruments for self-perceived knowledge. The relatively low difficulty level of questions might have limited the prevalence of misconceptions as these were more comprehensive than conceptual-oriented questions. Additionally, the post-test questions were not identical but isomorph to the pre-test questions which may have resulted in unanticipated changes in item difficulty because of near-transfer conversions of the original items. Furthermore, students had limited time to process the information as the total length of a seminar was approximately 120 min. An additional retention test might have provided more insight into the stableness of students’ conceptual understanding.

## Conclusion

This study was conducted to evaluate post-decision wagering as an instrument for measuring students’ self-perceived knowledge in educational contexts. Our findings add to the growing number of studies that demonstrate the subjective properties of PDW and thus do not support previous literature on the objectivity of PDW as a measure of confidence. In all, this study supports the use of Likert scales over post-decision wagering as an instrument for measuring self-perceived knowledge in educational settings.

**Conflict of interest** M. Versteeg and P. Steendijk declare that they have no competing interests.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



## References

- Ziegler B, Montplaisir L. Student Perceived and Determined Knowledge of Biology Concepts in an Upper-Level Biology Course. *CBE Life Sci Educ. IEEE Trans Med Imaging.* 2014;13:322-30.
- Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol.* 1999;77:1121-34.
- Flavell JH. Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *Am Psychol.* 1979;34:906.
- Zohar A, Barzilai S. A review of research on metacognition in science education: Current and future directions. *Stud Sci Educ. IEEE Trans Med Imaging.* 2013;49:121-69.
- Pintrich PR. The role of metacognitive knowledge in learning, teaching, and assessing. *Theory Pract.* 2002;41:219-25.
- Cavers M, Ling J. Confidence Weighting Procedures for Multiple-Choice Tests. In: Chen DG, Chen J, Lu X, Yi G, Yu H, editors. *Advanced Statistical Methods in Data Science. ICSA Book Series in Statistics.* Singapore: Springer; 2016.
- Sparck EM, Bjork EL, Bjork RA. On the learning benefits of confidence-weighted testing. *Cogn Res Princ Implic. IEEE Trans Med Imaging.* 2016;1:3.
- Swartz SM. Acceptance and Accuracy of Multiple Choice, Confidence-Level, and Essay Question Formats for Graduate Students. *J Educ Bus.* 2006;81:215-20.
- Hopkins KD, Hakstian AR, Hopkins BR. Validity and reliability consequences of confidence weighting. *Educ Psychol Meas.* 1973;33:135-41.
- Bokhorst FD. Confidence-weighting and the validity of achievement tests. *Psych Rep. IEEE Trans Med Imaging.* 1986;59:383-6.
- Brinkman DJ, Tichelaar J, van Agtmael MA, de Vries TP, Richir MC. Self-reported confidence in prescribing skills correlates poorly with assessed competence in fourth-year medical students. *J Clin Pharmacol.* 2015;55:825-30.
- Kampmeyer D, Matthes J, Herzog S. Lucky guess or knowledge: A cross-sectional study using the Bland and Altman analysis to compare confidence-based testing of pharmacological knowledge in 3rd and 5th year medical students. *Adv Health Sci Educ. IEEE Trans Med Imaging.* 2015;20:431-40.
- Rangel RH, Möller L, Sitter H, Stibane T, Strzelczyk A. Sure, or unsure? Measuring students' confidence and the potential impact on patient safety in multiple-choice questions. *Med Teach.* 2017;39:1189-94.
- Curtis DA, Lind SL, Boscardin CK, Dellenges M. Does student confidence on multiple-choice question assessments provide useful information? *Med Educ.* 2013;47:578-84.
- Lai NM, Teng CL. Self-perceived competence correlates poorly with objectively measured competence in evidence based medicine among medical students. *Bmc Med Educ.* 2011;11:25.
- Brydges R, Butler D. A reflective analysis of medical education research on self-regulation in learning and practice. *Med Educ.* 2012;46:71-9.
- Sandars J. The use of reflection in medical education: AMEE Guide. *Med Teach.* 2009;31(44):685-95.
- Leggett H, Sandars J, Burns P. Helping students to improve their academic performance: a pilot study of a workbook with self-monitoring exercises. *Med Teach.* 2012;34:751-3.
- Tanner K, Allen D. Approaches to biology teaching and learning: understanding the wrong answers—teaching toward conceptual change. *CBE Life Sci Educ. IEEE Trans Med Imaging.* 2005;4:112-7.
- Woods NN. Science is fundamental: the role of biomedical knowledge in clinical reasoning. *Med Educ.* 2007;41:1173-7.
- Dawson-Saunders B, Feltovich PJ, Coulson RL, Steward DE. A survey of medical school teachers to identify basic biomedical concepts medical students should understand. *Acad Med.* 1990;65:448-54.
- Michael JA, Wenderoth MP, Modell HI, et al. Undergraduates' understanding of cardiovascular phenomena. *Adv Physiol Educ.* 2002;26:72-84.
- Michael J. What makes physiology hard for students to learn? Results of a faculty survey. *Adv Physiol Educ.* 2007;31:34-40.
- Versteeg M, Wijnen-Meijer M, Steendijk P. Informing the uninformed: a multitier approach to uncover students' misconceptions on cardiovascular physiology. *Adv Physiol Educ.* 2018;43:7-14.
- Badenhorst E, Hartman N, Mamede S. How Biomedical Misconceptions May Arise and Affect Medical Students' Learning: A Review of Theoretical Perspectives and Empirical Evidence. *Health Prof Educ. IEEE Trans Med Imaging.* 2016;2:10:7.
- Posner GJ, Strike KA, Hewson PW, Gertzog WA. Accommodation of a scientific conception: Toward a theory of conceptual change. *Sci Educ. IEEE Trans Med Imaging.* 1982;66:211-27.
- Hasan S, Bagayoko D, Kelley EL. Misconceptions and the certainty of response index (CRI). *Phys Educ. IEEE Trans Med Imaging.* 1999;34:294-9.
- Gunstone RF, Mitchell IJ. Metacognition and conceptual change. In: Mintzes JJ, Wandersee JH, Novak JD, editors. *Teaching science for understanding. A human constructivist view.* San Diego, CA: Academic Press; 2005. pp. 133-63.
- Schraw G, Dennison RS. Assessing metacognitive awareness. *Contemp Educ Psychol.* 1994;19:460-75.
- Tippett CD. Refutation text in science education: A review of two decades of research. *Int J Sci Math Educ.* 2010;8:951-70.
- Lin LM, Zabrocky KM. Calibration of comprehension: Research and implications for education and instruction. *Contemp Educ Psychol.* 1998;23:345-91.
- Caleon IS, Subramaniam R. Do students know what they know and what they don't know? Using a four-tier diagnostic test to assess the nature of students' alternative conceptions. *Res Sci Educ. IEEE Trans Med Imaging.* 2010;40:313-37.
- Sreenivasulu B, Subramaniam R. Exploring undergraduates' understanding of transition metals chemistry with the use of cognitive and confidence measures. *Res Sci Educ. IEEE Trans Med Imaging.* 2014;44:801-28.
- Yan YK, Subramaniam R. Using a multi-tier diagnostic test to explore the nature of students' alternative conceptions on reaction kinetics. *Chem Educ Res Pract. IEEE Trans Med Imaging.* 2018;19:213-26.
- Brandriet AR, Bretz SL. The development of the redox concept inventory as a measure of students' symbolic and particulate redox understandings and confidence. *J Chem Educ.* 2014;91:1132-44.
- Koch C, Preuschoff K. Betting the house on consciousness. *Nat Neurosci.* 2007;10:140-1.
- Persaud N, McLeod P, Cowey A. Post-decision wagering objectively measures awareness. *Nat Neurosci.* 2007;10:257-61.
- Dienes Z, Scott R. Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychol Res.* 2005;69:338-51.
- Wierchoń M, Wronka E, Paulewicz B, Szczepanowski R. Post-decision wagering affects metacognitive awareness of emotional stimuli: an event related potential study. *PLoS ONE.* 2016;11:e159516.
- Sandberg K, Timmermans B, Overgaard M, Cleeremans A. Measuring consciousness: is one measure better than the other? *Conscious Cogn.* 2010;19:1069-78.
- Dienes Z, Seth A. Gambling on the unconscious: A comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. *Conscious Cogn.* 2010;19:674-81.



42. Fleming SM, Dolan RJ. Effects of loss aversion on post-decision wagering: implications for measures of awareness. *Conscious Cogn.* 2010;19:352–63.
43. Palizvan MR, Nejad MR, Jand A, Rafeie M. Cardiovascular physiology misconceptions and the potential of cardiovascular physiology teaching to alleviate these. *Med Teach.* 2013;35:454–8.
44. Lisk K, Agur AM, Woods NN. Exploring cognitive integration of basic science and its effect on diagnostic reasoning in novices. *Perspect Med Educ.* 2016;5:147–53.
45. Goldszmidt M, Minda JP, Devantier SL, Skye AL, Woods NN. Expanding the basic science debate: the role of physics knowledge in interpreting clinical findings. *Adv Health Sci Educ. IEEE Trans Med Imaging.* 2012;17:547:55.
46. Bell P, Volckmann D. Knowledge surveys in general chemistry: confidence, overconfidence, and performance. *J Chem Educ.* 2011;88:1469–76.
47. Favazzo L, Willford JD, Watson RM. Correlating student knowledge and confidence using a graded knowledge survey to assess student learning in a general microbiology classroom. *J Microbiol Biol Educ.* 2014;15:251.
48. Lindsey BA, Nagel ML. Do students know what they know? Exploring the accuracy of students' self-assessments. *Phys Rev Spec Top-Phys Educ R. IEEE Trans Med Imaging.* 2015;11:20103.
49. Potgieter M, Malatje E, Gaigher E, Venter E. Confidence versus performance as an indicator of the presence of alternative conceptions and inadequate problem-solving skills in mechanics. *Int J Sci Educ.* 2010;32:1407–29.
50. Ravesloot CJ, Van der Schaaf MF, Muijtjens AM, et al. The don't know option in progress testing. *Adv Health Sci Educ. IEEE Trans Med Imaging.* 2015;20:1325:38.
51. Anderson RC, Pearson PD. A schema-theoretic view of basic processes in reading comprehension. In: Pearson PD, Barr R, Kamil ML, Mosenthal P, editors. *Handbook of reading research.* New York: Longman; 1984. pp. 255–91.
52. Bransford JD, Johnson MK. Contextual prerequisites for understanding: Some investigations of comprehension and recall. *J Verbal Learn Verbal Behav.* 1972;11:717–26.
53. Alba JW, Hasher L. Is memory schematic? *Psychol Bull.* 1983;93:203.
54. Tse D, Langston RF, Kakeyama M, et al. Schemas and memory consolidation. *Science.* 2007;316:76–82.
55. van Kesteren MT, Ruiter DJ, Fernández G, Henson RN. How schema and novelty augment memory formation. *Trends Neurosci.* 2012;35:211–9.
56. Hewson MG, Hewson PW. Effect of instruction using students' prior knowledge and conceptual change strategies on science learning. *J Res Sci Teach.* 1983;20:731–43.
57. Hailikari T, Katajavuori N, Lindblom-Ylänne S. The relevance of prior knowledge in learning and instructional design. *Am J Pharm Educ.* 2008;72:113.

**Marjolein Versteeg** MSc (Hons), is a PhD student at Leiden University Medical Center, the Netherlands. She has a special interest in conceptual change and the use of neuroeducation in medical education research.

**Paul Steendijk** PhD, is a professor at Leiden University Medical Center, the Netherlands. He is a cardiovascular physiologist and actively involved in teaching medical physiology.