# The Incidence of Overconfidence and Underconfidence Effects in Medical Student Examinations

*Raúl A. Borracci, MD, and Eduardo B. Arribalzaga, MD*

Department of Surgery, School of Medicine, University of Buenos Aires, Buenos Aires, Argentina

**BACKGROUND:** Overconfidence is the tendency to overestimate the knowledge, capacity, or performance one really possesses. This cognitive bias could be potentially dangerous in medical decision-making, considering the impact it could have on patient health care. The aim of this study was to evaluate the incidence of overconfidence and underconfidence in medical student knowledge on general surgery by using traditional and new statistical approaches.

**METHODS:** During the application of a multiple-choice examination, 251 next-to-graduate medical students were invited to express the accuracy of their responses by choosing their own perceived confidence level for a set of questions. Analysis was done by comparing the difference between percentage of right answers (student's actual knowledge or accuracy) and self-estimated confidence level (student's perceived knowledge or confidence). Overconfidence was defined as a positive difference between confidence and accuracy, and underconfidence as a negative difference.

**RESULTS:** Nearly 12% of students showed significant overconfidence regarding their actual knowledge or accuracy levels. Better students showed a lower overconfidence effect than students with poorer performance. On the other hand, underconfidence was less likely than overconfidence (8.3% of students), and that effect was most frequently found in students who performed better in examinations.

**CONCLUSIONS:** The small proportion of our students exhibiting overconfidence or underconfidence behaviors moderates the need for educational interventions. Nevertheless, promoting prudence in individualized students manifesting overconfidence, and trust in those reporting significant underconfidence could increase the reliability of medical judgment during their future professional life. Overconfidence in individuals with lower scores in examinations may depend on a ceiling-like effect, since worst ranked students have a wider upper margin to manifest their confidence perceptions. The most confident students showed higher scores in examinations than the less confident ones. From this point of view, confidence could be considered an essential ingredient of success in examination performance. (J Surg Ed 75:1223-1229. ©2018 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

**KEY WORDS:** medical students, general surgery, education, cognitive sciences, trust

**COMPETENCIES:** Medical Knowledge

## INTRODUCTION

Charles Darwin wrote that ignorance often generates greater confidence than does knowledge.[1] More recently, the Dunning-Kruger effect was described, stating that the incompetent are often ill-suited to recognize their incompetence.[2,3] Originally, Kruger and Dunning[2] suggested that people who tend to overestimate their intellectual or social abilities suffer a dual burden: not only they reach erroneous conclusions and make unfortunate choices, but their incompetence robs them of the metacognitive ability to realize it. The overconfidence effect is a well-known bias in which a person's subjective confidence in his own judgments is reliably greater than the objective accuracy of those judgments.[4] Somehow, overconfidence is one example of miscalibration of subjective probabilities. Pallier et al.[5] have defined overconfidence in 3 different ways: (1) as an overestimation of one's actual performance, (2) as an overplacement of one's performance relative to others, and (3) as an excessive certainty regarding the accuracy of one's beliefs or knowledge, also termed overprecision.

Statement: The manuscript has been read and approved by all the authors, the requirements for authorship have been met, and each author believes that the manuscript represents honest work. The authors confirm all personal identifiers have been removed or disguised so the person(s) described are not identifiable and cannot be identified through the details of the story.

*Correspondence*: Inquiries to Raúl A. Borracci, Department of Surgery, School of Medicine, University of Buenos Aires, Av. Córdoba 2351, (C1120AAR), Buenos Aires, Argentina. Fax: +(54) 02322-48-2204; e-mail: raborracci@gmail.com

  1931-7204/$30.00 
https://doi.org/10.1016/j.jsurg.2018.01.015

Researchers in the field of cognitive psychology have shown that most people show an excessive level of confidence to answer a series of questions on general knowledge, after asking them to express the accuracy of their responses.[6,7] By analogy, overconfidence could be potentially dangerous in medical decision-making, considering the direct impact it could have on patient health care.[8] Berner et al.[9] argued that physicians in general underappreciate the likelihood that their diagnoses may be wrong.

While doctors need to demonstrate a certain level of confidence in their skills to interact with patients, they also need to know the wrong foundations of that confidence. Academic tests measured by a confidence-weighted scoring technique showed that overconfidence and underconfidence indices help to predict medical students' subsequent academic achievements.[10] Other investigations found that students tended toward underconfidence in their diagnostic judgments when classifying heart arrhythmias.[11] Furthermore, nearly 19% of medical residents showed to be overconfident when diagnosing complex clinical cases,[12] while faculty physicians were overconfident in 13% of diagnoses.[13] Cross-cultural variations in probability judgment accuracy and confidence were also recognized, even in medical diagnoses.[14] Some concerns exist about some counterintuitive effects of confidence (hard-easy and underconfidence-with-practice effects), and about the correspondence between subjective and objective probabilities, since the same data can appear to reveal both overconfidence and underconfidence, depending on the method of data analysis.[15,16] According to these observations, other authors proposed that somehow, confidence level could be considered a statistical artifact.[17,18]

Based on this theoretical framework we hypothesized that next-to-graduate medical students could show a higher or lower level of confidence than actually justified by their knowledge or performance. Therefore, the aim of this work was to evaluate the incidence of overconfidence and underconfidence in medical student knowledge by using traditional and new statistical approaches.

## MATERIAL AND METHODS

From May 2015 to September 2016, a prospective study was conducted at the Buenos Aires University School of Medicine. While taking a multiple-choice examination of general surgery, 251 next-to-graduate medical students (Fifth-year students in Argentina) expressed the accuracy of their responses by choosing their own perceived confidence level for each question on a 5-point scale (20, 40, 60, 80, and 100). Each multiple-choice question had 5 options, and only 1 right answer. After selecting the presumed right answer, students pointed out the confidence level they estimated for the response. The procedure was repeated for a total of 60 questions, in order to obtain the percentage of right answers and the average percentage of confidence levels selected by each student for the whole examination. Statistical analysis was done by comparing the difference between percentage of right answers (student's actual knowledge or accuracy) and self-estimated confidence level (student's perceived knowledge or confidence). Overconfidence (that is to say, overprecision) was operatively defined as a positive difference between confidence and accuracy, and underconfidence (that is to say, underprecision) as a negative difference between those values. Statistically significant values of overconfidence and underconfidence were considered when a particular confidence-to-accuracy difference exceeded the 95% confidence limit.

### Statistical Analysis

Continuous variables were expressed as mean and standard deviation or 95% CI, or median and 25-75 percentile ($P_{25\%-75\%}$). Kolmogorov-Smirnov goodness-of-fit test was used to analyze normal distributions. Independent metric variables were compared with Student *t*-test, and paired variables (concordance/accuracy) with paired Student *t*-test. Comparison of dichotomous variables was performed using the $\chi^2$ test and the odds ratio (OR) with the associated 95% CI. Two-tailed Fisher exact test was employed when cell expected values were ≤5. A modified (nonbinary) mean probability score, or Brier score, was used to assess the statistical consistency between students' confidence and accuracy, and calibration was determined with Spiegelhalter's *Z*-statistical method.[19] The Brier score is useful to measure the predictive accuracy of a judgment. It compares the observed probability ($y$) (in the current study, the percentage of right answers) with the expected probability (p) (the mean confidence level) for each student ($i$), according to the following equation:

$\text{Brier} = 1/n \; \Sigma \; [(y_i - p_i)/100]^2$, where $n$ is the number of students. Nonparametric Kendall $\tau_b$ was used as a measure of association to test the significance of the confidence/accuracy relationship. The output of the expected-to-observed (E-O) (confidence-to-accuracy) percentage of right answers for each student was calculated and plotted. The plot $y$-axes computed the expected minus the observed percentage of right answers. The expected percentage corresponds to the confidence level as estimated by each student, and the observed percentage indicates the accuracy or proportion of right answers obtained in the examination. To determine statistically significant overconfidence or underconfidence, 95% confidence limits of E-O difference were calculated for each student's performance with the confidence interval of the difference between 2 proportions. In addition, the statistical analyses were performed with SPSS Statistics for Windows, Version 17.0. Chicago: SPSS

Inc. and a 2-tailed $p \leq 0.05$ was considered statistically significant.

## Ethical Considerations

Confidentiality was ensured to participants responding to the questionnaire. All respondents voluntarily participated in the study after being explained its purpose, and expressed consent by filling the form. All personal identifiers were removed or disguised so the students and physicians described were not identifiable and could not be identified through the details of the study. Heads of medical training institutions provided access to the student population after ethical approval of the protocol. Ethical clearance for this study was granted by the Ethics Committee of the Buenos Aires University School of Medicine.

## RESULTS

From a total of 251 students surveyed, 228 (90.8%) completed the test. Mean student age was 26.7 years (standard deviation is 1.84), and 64.5% were females (*n*: 147). The cohort of female respondents was proportional to the number of female medical students in the country. Average accuracy level was 79.0% (95% CI: 78.0-80.1), while mean confidence level was 79.7% (95% CI: 78.3-81.0) (paired Student *t*-test $p = 0.417$). The mean Brier score was 0.013 (median 0.005, $P_{25\%-75\%}$ 0.001-0.018) for the overall cohort, showing good predictive accuracy but poor calibration according to Spiegelhalter's method. ($Z$ 21.4, $p < 0.0001$). Boxplot of Brier's score *nongaussian* distribution was plotted in Figure 1, where outliers show students with significant overconfidence or underconfidence. According to this approach, 6.1% (*n*: 14) of students were shown to be overconfidence or underconfident regarding their accuracy levels. Kendall $\tau_b$ measure of association between accuracy and confidence showed a small but statistically significant value ($\tau_b$ 0.202, $p < 0.0001$). Since the Brier score equation is squared, it was not possible to discriminate between overconfidence and underconfidence (positive or negative differences) with this method. To overcome the limitations of this method and maintain the positive and negative differences between confidence and accuracy, the output of the E-O difference for each student was used. Based on E-O difference, 27 students among the total cohort (11.8%) significantly overestimated their accuracy levels, while 19 students (8.3%) significantly underestimated their accuracy levels. Gender differences with regard to overconfidence were 12.4% vs 11.6% for males and females, respectively ($p = 0.861$), and 8.6% vs 8.2% in the case of underconfidence ($p = 0.900$). Figure 2 shows the E-O (confidence-to-accuracy) percentage of right answers in the examination for each student. The *y*-axis computes the expected minus the observed (E-O) percentage of right answers. The expected percentage corresponds to the confidence level as estimated by each student, and the observed percentage indicates the proportion of right answers obtained in the examination

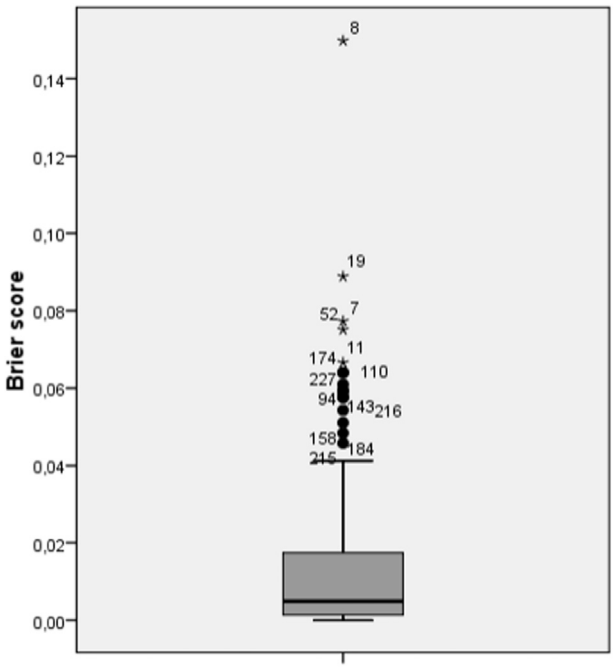


**FIGURE 1.** Boxplot of Brier scores with a positive asymmetric distribution. Outliers show students with significant misplaced confidence. Since Brier score values are always positive, it is not possible to discriminate between overconfidence and underconfidence cases.

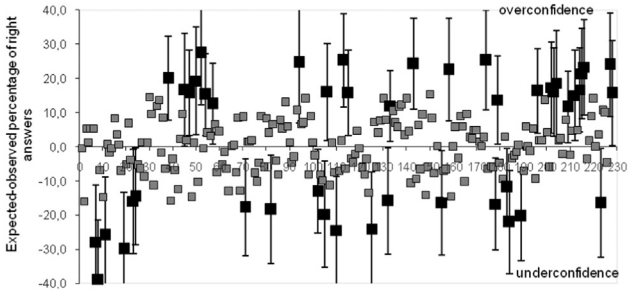


**FIGURE 2.** The output of the expected-to-observed (confidence-to-accuracy) percentage of right answers for each student. The *y*-axes computes the expected minus the observed (E-O) percentage of right answers. The expected percentage corresponds to the confidence level as estimated by each student, and observed percentage indicates the proportion of right answers obtained in the examination. Little grey squares show individuals taking the examination with a not significant E-O difference; while larger black squares indicate those individuals with a statistically significant E-O difference ($p < 0.05$). Vertical lines on black squares correspond to the 95% CI (nonsignificant intervals crossing the E-O zero value were eliminated). Individuals in the lowest bottom half of the chart had a lower estimated confidence level than the true percentage of right answers obtained in the examination. Individuals in the highest top half of the panel had a higher estimated confidence level than the true percentage of right answers obtained in the examination. It was less probable for students to have a lower confidence level (underconfidence) than a higher 1 (overconfidence), in comparison to the actual number of right responses.

(accuracy). Small gray squares in the plot show individuals taking the examination with a not significant E-O difference, while larger black squares indicate those individuals with a statistically significant E-O difference ($p < 0.05$). Individuals in the lowest bottom half of the chart had a lower estimated confidence level than the true percentage of right answers obtained in the examination, while individuals in the highest top half of the panel had a higher estimated confidence level than the true percentage of accuracy.

When the series was divided into 2 groups, over and under the median value of accuracy, the average positive differences between confidence and accuracy (overconfidence) were 4.9% (95% CI: 3.8-6.0) for the students with better examination performance (those over the median value of accuracy), and 10.9% (95% CI: 9.3-12.5) for the students with poorer performance (those under the median value of accuracy) ($p < 0.0001$). Among students with poorer performance, 21.1% (*n*: 24) showed significant overconfidence, vs 2.6% (*n*: 3) of better students (OR = 9.9, 95% CI: 2.9-33.8; $p < 0.0001$). On the other hand, the mean negative difference between confidence and accuracy (underconfidence) was 9.3% (95% CI: 7.4-11.2) and 8.6% (95% CI: 6.7-10.6) for students with better and poorer performance, respectively ($p = 0.662$). In this case, 15.8% (*n*: 18) of the best students showed significant underconfidence, vs 0.9% (*n*: 1) of the students with poorer performance (OR = 21.2, 95% CI: 2.8-161.6; $p < 0.0001$). Finally, when the series was divided into 2 groups according to the median confidence level (81.3%), students in the top half had an average accuracy of 81.0% (95% CI: 79.7-82.4) vs 77.0 % (95% CI: 75.5-78.6) for students under the median confidence level ($p = 0.0002$).

## DISCUSSION

Physician overconfidence is considered one of the many factors contributing to medical errors, and occurs when the relationship between accuracy and confidence is miscalibrated, such that confidence is higher than it should be.[20-22] In this study, nearly 12% of students showed a significant overconfidence effect regarding their actual knowledge or accuracy levels. However, students with better examination performance showed a lower overconfidence effect than students with poorer performance, when either raw data or average differences between confidence and accuracy were considered. On the other hand, underconfidence was less likely than overconfidence (nearly 8% of students), and that effect was most frequently found in students with better than with poorer performance when using raw data analysis.

Individuals with higher scores in examination tended to show lower overconfidence level; on the contrary, students with lower scores showed higher degree of overconfidence. These outcomes may depend on a ceiling-like effect, since best ranked students have a limited top margin to express a confidence level over their percentage of right answers; on the other hand, worst ranked individuals have a wider margin to manifest their confidence perceptions. To avoid this statistical artifact, the relationship between performance and confidence may be better assessed by considering the underconfidence level. Some authors recognized that students with poorer performance showed a greater overconfidence effect, but they also reported lower confidence in their own predictions, showing that indeed unskilled students had some awareness of their lack of knowledge.[23] In a previous research, Langendyk[24] evaluated the accuracy of self- and peer-assessment in 175 third-year students according to their academic performance. The cohort-based analysis suggested that the majority of the students could judge accurately the quality of their own performance and that of their peers; but low-achieving students were less likely to be able to assess accurately their own. In our study, students with poorer performance were only related with overconfidence, while better students were more associated with underconfidence. A student with a poor performance had nearly 10 times more chances of being overconfident than a better student; and on the contrary, a student with a good performance had nearly 21 times more chances of being underconfident than a student with poorer performance. It is possible that students adjusted their responses relative to the cost of overestimating or underestimating the correct answer, since they firstly wanted to have a good performance in the examination.[25]

In contrast, the alternative analysis with the Brier score revealed a combined overconfidence and underconfidence rate of only 6.1%. Though the statistical analysis of confidence data has been often based on the Brier score, its complementary Spiegelhalter calibration method has been rarely used. In our study, a mean Brier score near zero showed a good predictive accuracy, but a poor calibration index. Though the overconfidence and underconfidence rates we found vary depending on the statistical analysis approach, these proportions are significantly lower than those reported by other authors with a dichotomized analysis of medical students, ranging from 18% to 25% of overconfidence, and 16% of underconfidence.[12,13]

Confidence is thought to be an essential ingredient of success in job performance, and even overconfidence seems to increase the probability of individual success.[26] Nevertheless, overconfidence also leads to unrealistic expectations and hazardous decisions, compromising a safe and accurate medical practice. Conversely, underconfidence could induce physicians to deliberate further or seek additional diagnostic help before decision-making. Previous research found that highly overconfident students rank better in medical examinations than slightly overconfident ones.[9] In our study, we confirmed that the more confident students had a better performance in the examination than the less confident individuals.

Machiavellianism (Mach) in medicine is a trait reflecting an individual's pursuit of self-interest at any expense, which can also involve overconfidence. Although males tend to

present higher Mach scores than females, gender-based Mach scoring can vary significantly among populations of medical students.[27-29] In this study, no Mach scale was used to assess students' perception of their own confidence. Moreover, no gender differences were found with regard to overconfidence, though the sample size could be underpowered to detect statistical differences. In a classical study, Ehrlinger and Dunning[30] demonstrated that an important source of people's perception of their own performance are deep-seated views they hold regarding their abilities. Particularly, these authors observed that although women performed equally to men on a science quiz, yet they tended to underestimate their performance. Again, no gender differences were found with regard to underconfidence in the current study.

Several questions on the topic remain controversial. The confidence level seems to be relatively independent of both diagnostic accuracy and case complexity when physicians are challenged to resolve clinical cases of increasing difficulty.[31] Other authors have also reported that overconfidence and underconfidence vary systematically with the domain of questions asked, but not as a function of difficulty.[32] In contrast, many other studies have indicated that overconfidence increases with the difficulty of the task.[33-35] This last position may be supported by the sequence *more difficulty/less accuracy/more confidence*. Furthermore, the validity of peer-comparison questions employed to measure overconfidence has been also questioned from a methodological viewpoint.[36-38]

Ehrlinger et al.[39] observed that good students became more accurate in predicting how they would do on future examinations; but the poorest performers did not, showing no recognition despite repeated feedback. Consequently, they continued to provide overly optimistic predictions about how well they would do in future tests. These findings raise some concern about the possibility of implementing some intervention to reduce this cognitive bias.

Some limitations of this study should be acknowledged. It is important to consider a possible bias toward overconfidence, since a statistical effect favoring overconfidence is inevitable if student judgments are imperfect and examination responses are wrong. Probably, students' confidence level would vary if they confronted a real clinical situation than a paper-based multiple-choice examination. At present, there is no conclusive evidence on the effect of improving the realism of clinical judgment tasks on overconfidence and underconfidence.[40] Another limitation is that demonstrating misplaced confidence among students does not necessarily mean it entails consequences or benefits, or that these biases are necessarily a problem.[41]

## CONCLUSIONS

In this study, nearly 12% of medical students were overconfident regarding their actual knowledge or accuracy levels, while nearly 8% were underconfident. Since these proportions were significantly lower than those reported by other authors, the small percentage of our students exhibiting overconfidence or underconfidence behaviors curbs the need for educational interventions. Nevertheless, promoting prudence in individualized students manifesting overconfidence, and trust in those reporting significant underconfidence could increase the reliability of medical judgment during their future professional life. Notwithstanding the previously recognized observation that students with higher scores in examinations tended to be underconfident, and individuals with lower scores were generally overconfident, overconfidence in the last group may depend on a ceiling-like effect, since worst ranked students have a wider upper margin to manifest their confidence perceptions than better ranked individuals. Finally, although no cause-effect relationship was established, in this study the most confident students showed higher scores in examinations than the less confident ones. From this point of view, confidence could be considered an essential ingredient of success in examination performance.

## REFERENCES


1. Darwin C. The Descent of Man. London: John Murray; 1874.
2. 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(6):1121-1134.
3. Pennycook G, Ross RM, Koehler DJ, Fugelsang JA. Dunning-Kruger effects in reasoning: theoretical implications of the failure to recognize incompetence. *Psychon Bull Rev*. 2017;24(6):1774-1784.
4. Pallier G, Wilkinson R, Danthiir V, et al. The role of individual differences in the accuracy of confidence judgments. *J Gen Psychol*. 2002;129(3):257-299.
5. Moore DA, Healy OJ. The trouble with overconfidence. *Psychol Rev*. 2008;115(2):502-517.
6. Lichtenstein S, Fischhoff B, Phillips LD. Calibration of probabilities: the state of the art to 1980. Kahneman Daniel, Slovic Paul, Tversky Amos, editors. Judgment Under Uncertainty: Heuristics and Biases. Cambridge, UK: Cambridge University Press, 1982, pp. 306-334 ISBN 978-0-521-28414-1.
7. Chabris CF, Schuldt J, Wolley AW. Individual differences in confidence affect judgments made collectively by groups. May 25-8. New York: Annual Convention of the Association for Psychological Science; 2006.

8. Hodges B, Regehr G, Martin D. Difficulties in recognizing one′s own incompetence: novice physicians who are unskilled and unaware of it. *Acad Med*. 2001;76(suppl 10):S87-S89.

9. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med*. 2008;121 (suppl 5):S2-S23.

10. Zeleznik C, Hojat M, Goepp CE, Amadio P, Kowlessar OD, Borenstein B. Students′ certainty during course test-taking and performance on clerkships and board exams. *J Med Educ*. 1988;63(12): 881-891.

11. Mann D. The relationship between diagnostic accuracy and confidence in medical students. Atlanta. Paper presented at the Annual Meeting of the American Educational Research Association; 1993.

12. Friedman C, Gatti G, Elstein A, Franz T, Murphy G, Wolf F. Are clinicians correct when they believe they are correct? Implications for medical decision support London: Proceedings of the Tenth World Congress on Medical Informatics; 2001;10(PP.1):454-458.

13. Friedman CP, Gatti GG, Franz TM, et al. Do physicians know when their diagnoses are correct? Implications for decision support and error reduction *J Gen Intern Med*. 2005;20(4):334-339.

14. Yates JF, Lee JW, Shinotsuka H, Patalano AL, Sieck WR. Cross-cultural variations in probability judgment accuracy: beyond general knowledge overconfidence? *Organ Behav Hum Decis Process*. 1998;74(2):89-117.

15. Merkle EC. The disutility of the hard-easy effect in choice confidence. *Psychon Bull Rev*. 2009;16(1): 204-213.

16. Koriat A, Sheffer L, Ma'ayan H. Comparing objective and subjective learning curves: judgments of learning exhibit increased underconfidence with practice. *J Exp Psychol Gen*. 2002;131(2):147-162.

17. (a) Erev I, Wallsten TS, Budescu DV. Simultaneous over- and underconfidence: the role of error in judgment processes. *Psychol Rev*. 1994;101(3):519-527; (b) Ayton P, Mc Clelland AGR. How real is overconfidence? *J Behav Decis Mak*. 1997;10:279-286.

18. Brenner L. Should observed overconfidence be dismissed as a statistical artifact? Critique of Erev, Wallsten, and Budescu (1994) *Psychol Rev*. 2000;107 (4):943-946.

19. Spiegelhalter DJ. Probabilistic prediction in patient management and clinical trials. *Stat Med*. 1986;5(5): 421-433.

20. Pusic MV, Boutis K, Hatala R, Cook DA. Learning curves in health professions education. *Acad Med*. 2015;90(8):1034-1042.

21. Croskerry P, Norman G. Overconfidence in clinical decision making. *Am J Med*. 2008;121(5A):S24-S29.

22. Price PC, Stone ER. Intuitive evaluation of likelihood judgment producers: evidence for a confidence heuristic. *J Behav Decis Mak*. 2004;17:39-57.

23. Miller TM, Geraci L. Unskilled but aware: reinterpreting overconfidence in low-performing students. *J Exp Psychol Learn Mem Cogn*. 2011;37(2):502-506.

24. Langendyk V. Not knowing that they do not know: self-assessment accuracy of third-year medical students. *Med Educ*. 2006;40(2):173-179.

25. Mannes AE, Moore DA. A behavioral demonstration of overconfidence in judgment. *Psychol Sci*. 2013;24(7): 1190-1197.

26. Johnson DD, Fowler JH. The evolution of overconfidence. *Nature*. 2011;477(7364):317-320.

27. Garcia D, Rosenberg P, MacDonald S, Räisänen C, Ricciardi MR. Measuring malevolent character: data using the Swedish version of Jonason′s dark triad dirty dozen. *Data Brief*. 2017;14:648-652.

28. Hren D, Vujaklija A, Ivanisević R, Knezević J, Marusić M, Marusić A. Students′ moral reasoning, Machiavellianism and socially desirable responding: implications for teaching ethics and research integrity. *Med Educ*. 2006;40(3):269-277.

29. Bratek A, Bonk M, Bulska W, Tyrała K, Seweryn M, Krysta K. Claw your way–Machiavellianism among the medical community. *Psychiatr Danub*. 2015;27 (Suppl 1):S344-S347.

30. Ehrlinger J, Dunning D. How chronic self-views influence (and potentially mislead) estimates of performance. *J Pers Soc Psychol*. 2003;84(1):5-17.

31. Meyer AN, Payne VL, Meeks DW, Rao R, Singh H. Physicians' diagnostic accuracy, confidence, and resource requests: a vignette study. *J Am Med Assoc Intern Med*. 2013;173(21):1952-1658.

32. Klayman J, Soll JB, González-Vallejo C, Barlas S. Overconfidence: it depends on how, what, and whom you ask. *Organ Behav Hum Decis Process*. 1999;79(3): 216-247.

33. Soll JB. Determinants of overconfidence and miscalibration: the roles of random error and ecological structure. *Organ Behav Hum Decis Process*. 1996;65(2): 117-137.

**34.** Suantak L, Bolger F, Ferrell WR. The hard-easy effect in subjective probability calibration. *Organ Behav Hum Decis Process*. 1996;67(2):201-221.

**35.** Baranski JV, Petrusic WM. The calibration and resolution of confidence in perceptual judgments. *Percept Psychophys*. 1994;55(4):412-428.

**36.** Colvin CR, Block J. Do positive illusions foster mental health? An examination of the Taylor and Brown formulation *Psychol Bull*. 1994;116(1):3-20.

**37.** Gramzow RH, Elliot AJ, Asher E, McGregor HA. Self-evaluation bias and academic performance: some ways and some reasons why. *J Res Pers*. 2003;37(2):41-61.

**38.** Bi YL, Du XL, Li S. Peer-comparison overconfidence: does it measure bias in self-evaluation? *Psych J*. 2013;2(1):17-25.

**39.** Ehrlinger J, Johnson K, Banner M, Dunning D, Kruger J. Why the unskilled are unaware: further explorations of (absent) self-insight among the incompetent. *Organ Behav Hum Decis Process*. 2008;105(1):98-121.

**40.** Yang H, Thompson C, Bland M. Effect of improving the realism of simulated clinical judgment tasks on nurses′ overconfidence and underconfidence: evidence from a comparative confidence calibration analysis. *Int J Nurs Stud*. 2012;49(12):1505-1511.

**41.** Morgan PJ, Cleave-Hogg D. Comparison between medical students' experience, confidence and competence. *Med Educ*. 2002;36(6):534-539.