

ORIGINAL RESEARCH

Education

The Dunning–Kruger effect in resident predicted and actual performance on the American Board of Emergency Medicine in-training examination

Theodore J. Gaeta DO, MPH¹ | Earl Reisdorff MD² | Melissa Barton MD² |
Kim M. Feldhaus MD³ | Marianne Gausche-Hill MD⁴ | Deepi Goyal MD⁵ |
Kevin Joldersma PhD² | Chadd K. Kraus DO, DrPH² | Felix Ankel MD⁶

¹Department of Emergency Medicine, New York Presbyterian Brooklyn Methodist Hospital, Brooklyn, New York, USA

²American Board of Emergency Medicine, East Lansing, Michigan, USA

³Department of Emergency Medicine, Boulder Community Health, Boulder, Colorado, USA

⁴Departments of Emergency Medicine and Pediatrics, Harbor-UCLA Medical Center, Los Angeles, California, USA

⁵Department of Emergency Medicine, Mayo Clinic, Rochester, Minnesota, USA

⁶Department of Emergency Medicine, Regions Hospital, University of Minnesota Medical School, St. Paul, Minnesota, USA

Correspondence

Theodore J. Gaeta, Department of Emergency Medicine, New York Presbyterian Brooklyn Methodist Hospital, Brooklyn, NY, USA.
Email: ted.gaeta@gmail.com

Prior presentation: 2023 SAEM Annual Meeting, Austin, Texas, USA.

Abstract

Objectives: The Dunning–Kruger effect (DKE) is a cognitive bias wherein individuals who are unskilled overestimate their abilities, while those who are skilled tend to underestimate their capabilities. The purpose of this investigation is to determine if the DKE exists among American Board of Emergency Medicine (ABEM) in-training examination (ITE) participants.

Methods: This is a prospective, cross-sectional survey of residents in Accreditation Council for Graduate Medical Education (ACGME)-accredited emergency medicine (EM) residency programs. All residents who took the 2022 ABEM ITE were eligible for inclusion. Residents from international programs, residents in combined training programs, and those who did not complete the voluntary post-ITE survey were excluded. Half of the residents taking the ITE were asked to predict their self-assessment of performance (percent correct), and the other half were asked to predict their performance relative to peers at the same level of training (quintile estimate). Pearson's correlation (r) was used for parametric interval data comparisons and a Spearman's coefficient (ρ) was determined for quintile-to-quintile comparisons.

Results: A total of 7568 of 8918 (84.9%) residents completed their assigned survey question. A total of 3694 residents completed self-assessment (mean predicted percentage correct 67.4% and actual 74.6%), with a strong positive correlation (Pearson's r 0.58, $p < 0.001$). There was also a strong positive correlation (Spearman's ρ 0.53, $p < 0.001$) for the 3874 residents who predicted their performance compared to peers. Of these, 8.5% of residents in the first (lowest) quintile and 15.7% of residents in the fifth (highest) quintile correctly predicted their performance compared to peers.

Conclusions: EM residents demonstrated accurate self-assessment of their performance on the ABEM ITE; however, the DKE was present when comparing their

Supervising editor: Katherine Edmunds, MD, Med

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *Journal of the American College of Emergency Physicians Open* published by Wiley Periodicals LLC on behalf of American College of Emergency Physicians.

self-assessments to their peers. Lower-performing residents tended to overestimate their performance, with the most significant DKE observed among the lowest-performing residents. The highest-performing residents tended to underestimate their relative performance.

1 | INTRODUCTION

1.1 | Background

Self-assessment and directed learning play pivotal roles in the development and maintenance of clinical proficiency.¹ Studies have revealed that physicians, in general, possess a limited capacity for accurately assessing their own abilities.² This holds true for emergency medicine (EM) residents, who exhibit varying abilities in self-assessment.³

The Dunning–Kruger effect (DKE) is a cognitive bias wherein individuals who are unskilled overestimate their abilities, while those who are skilled tend to underestimate their capabilities.⁴ Interestingly, the more confident one becomes in their abilities, the more pronounced this overestimation of knowledge and skills becomes. Moreover, people often struggle to accurately gauge how their abilities compare to those of others. In particular, individuals with lower skills frequently lack metacognitive insight into their relative performance and disproportionately rate their self-assessments as superior.^{4–6}

Metacognition, defined as the ability to think about one's own thinking, consists of two key elements: metacognitive knowledge and metacognitive regulation. Metacognitive knowledge includes learning processes, awareness of effective learning strategies, and the ability to distinguish between knowing and not knowing. Metacognitive regulation refers to learners' ability to accurately evaluate strengths and weaknesses, reflect on the success of their strategies, and adjust accordingly. One method used to assess metacognitive monitoring ability includes the calculation of the difference between self-predictive and actual performance. Both over- and underestimates lead to misalignment in one's judgment between perceived and actual performance.⁷

1.2 | Importance

Emergency physicians routinely make critical decisions during their shifts, often under high-stakes circumstances. Clinical reasoning and decision-making processes are complex and prone to errors. Cognitive error can also be difficult to identify and is equally difficult to prevent. To mitigate patient harm resulting from errors in critical thinking, it has been proposed to train physicians to understand and maintain awareness of their thought process, to identify error-prone clinical situations, to recognize predictable vulnerabilities in thinking, and to use strategies to avert cognitive errors.⁸ DKE arises from the following tendencies: (1) overestimation of one's own skill level; (2) underestimation of others' skill and expertise of others; and (3) failing to recognize one's

own mistakes and lack of skill. Recognition and understanding of these cognitive biases are important for trainees in graduate medical education to ensure the highest level of lifelong learning habits based on accurate self-assessment.⁹ Identification of DKE in EM residents might provide program leadership an opportunity for using supervision and feedback as countermeasures.¹⁰

1.3 | Goals of the investigation

The objective of this study is to determine if the DKE exists among residents taking the American Board of Emergency Medicine (ABEM) in-training examination (ITE) in their self-estimation of performance and assessment of performance relative to their peers. This information can help residency training programs focus on educational interventions and professional development opportunities to mitigate the impact of overconfidence among EM residents.

2 | METHODS

2.1 | Study setting and population

We conducted a cross-sectional study of EM residents in categorical, Accreditation Council for Graduate Medical Education (ACGME)-accredited EM residency programs who completed the ABEM ITE and a voluntary survey at the conclusion of the examination. The examination was administered from February 22 to 28, 2022. Participants were provided a statement before the survey that informed them of the survey's purpose and that it was voluntary, all data would be deidentified, program directors and department chairs would not have access to any responses, and participation would have no effect on the results of any ABEM examination. Demographic data were self-reported.

EM residents who took the 2022 ABEM ITE were eligible for inclusion. Residents from international programs, residents in combined training programs (e.g., internal medicine–EM), and those who did not complete the voluntary post-ITE survey were excluded from the study.

2.2 | Study protocol

Residents were randomly assigned to one of two survey questions (A or B). Each survey contained a different question to assess possible DKE: question A asked them to predict their absolute percentage correct on the ITE, and question B asked them to predict their performance quintile relative to peers at the same level of training. Responses were

compared to their ITE score, estimated percentile of performance, or quintile of performance on the February 2022 ITE (Appendix 1—Survey questions A and B). The large cohort allowed us to investigate two independent means of evaluating DKE among EM residents. First, self-assessment compared to absolute objective performance (question A), and second, self-assessment compared to relative performance (question B). Both indicators of DKE have been previously studied.^{11,12}

This study was deemed to be exempt for human research by the NYP Brooklyn Methodist Hospital Institutional Review Board.

2.3 | Key outcome measures

The ITE test score is a scaled score that is an equated count of scorable items transformed to a scale ranging from 0 to 100. We used the EM year of training for “level of training” to compare resident ITE scores against other residents at the same level. We generated percentiles and quintiles for examination performance relative to peers.

2.4 | Data analysis

Pearson's correlation (r) was used for parametric interval data comparisons and Spearman's coefficient (ρ) was determined for quintile-to-quintile comparisons. All other analyses are frequencies and descriptive statistics of the score distribution and survey responses. Data preparation and analysis were conducted in Stata¹³ and verified using R.¹⁴

3 | RESULTS

Of 8918 EM residents in 272 ACGME-accredited EM residencies who took the ABEM ITE, 7568 (84.9%) completed a post-ITE survey.

The Bottom Line

A post-exam survey was completed by 7,568 emergency medicine residents who took the 2022 American Board of Emergency Medicine (ABEM) in-training examination (ITE). They were asked to estimate their own performance on the ITE and compare it to their peers. Residents were accurate in predicting their own performance on the ITE (Pearson $r = 0.58$, $p < 0.001$). However, lower-performing residents tended to overestimate their performance relative to peers, and the highest-performing residents tended to underestimate their performance (Spearman $\rho = 0.53$, $p < 0.001$). The mismatch in self-awareness suggests an opportunity for training programs and certification bodies to develop strategies that promote accurate self-assessment and encourage continuous learning.

The responders included 4579 males (60.5%), 2982 females (39.4%), 4728 non-Hispanic White residents (62.5%) and 2840 residents from other racial/ethnic groups with comparable demographic distribution between the two survey formats and between responders and non-responders (Table 1).

A total of 3694 residents provided responses to question A, where they predicted the percentage of correct answers. In this group, the predicted mean was 67.4%, while the actual performance showed a mean scaled score of 74.6%. There was a strong positive correlation between absolute predicted score and the actual obtained score (Pearson's $r = 0.58$, $p < 0.001$) (Figure 1). Residents underpredicted their performance across the ability spectrum with visibly greater inaccuracy at the highest quintile.

TABLE 1 Race/ethnicity by survey form completed.

Race/ethnicity	Survey A		Survey B		Total		Non-responders	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Asian or Pacific Islander	508	13.8	550	14.2	1,058	14.0	212	15.7
Alaska Native	2	0.1	1	0.0	3	0.0	1	0.1
Black (non-Hispanic)	188	5.1	204	5.3	392	5.2	79	5.9
Other Hispanics	151	4.1	176	4.5	327	4.3	53	3.9
American Indian	19	0.5	16	0.4	35	0.5	5	0.4
Mexican American	44	1.2	56	1.4	100	1.3	16	1.2
Native American	20	0.5	30	0.8	50	0.7	8	0.6
Native Hawaiian	2	0.1	4	0.1	6	0.1	2	0.1
Other Hispanics	134	3.6	132	3.4	266	3.5	83	6.1
Puerto Rican	36	1.0	38	1.0	74	1.0	10	0.7
Unknown	266	7.2	263	6.8	529	7.0	93	6.9
White	2324	62.9	2404	62.1	4728	62.5	788	58.4
Total	3694	100.0	3874	100.0	7568	100.0	1350	100.0

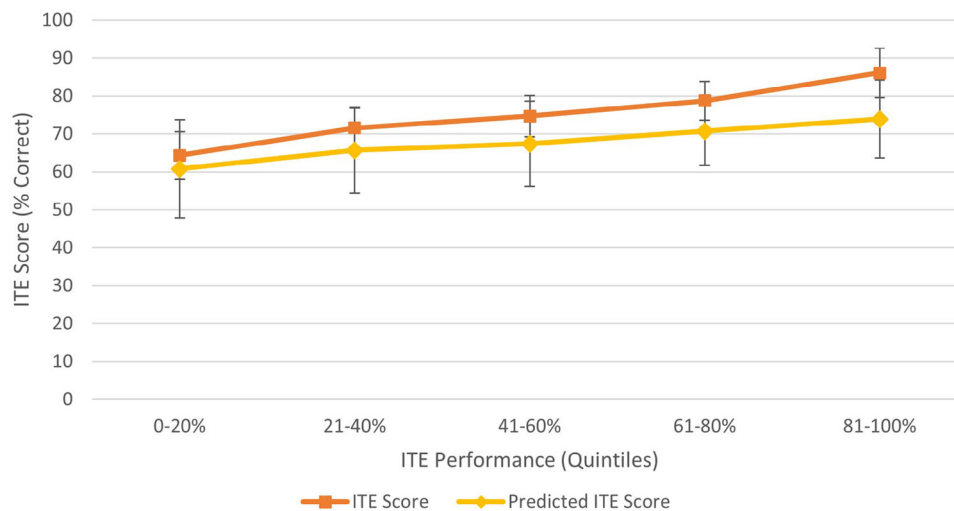


FIGURE 1 In-training examination (ITE) score and predicted percent correct by overall quintiles.

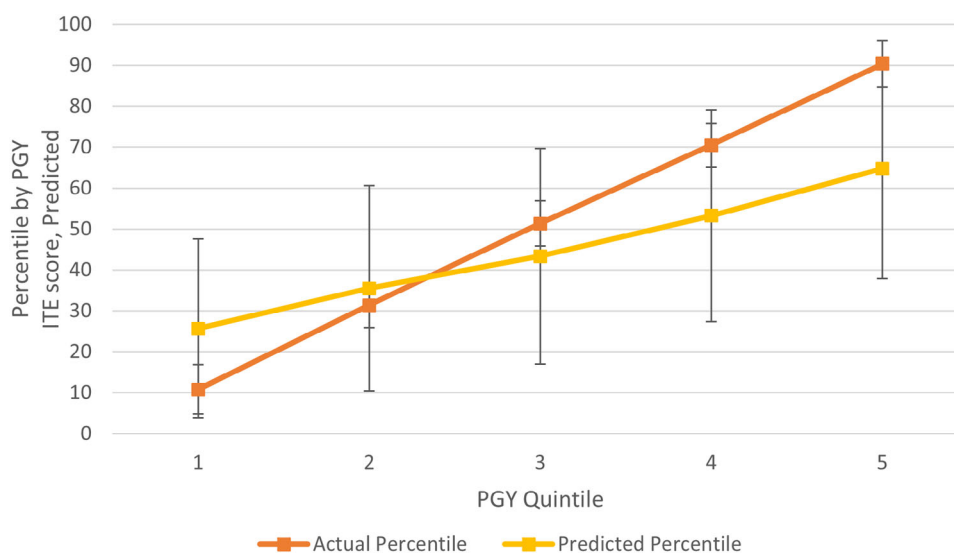


FIGURE 2 In-training examination (ITE) score as a percentile and predicted percent correct as a percentile within Post Graduate Year (PGY) quintiles.

Question B asked residents to estimate their performance relative to their peers and was completed by 3874 residents. There was also a strong positive correlation between perceived relative quintile and actual quintile of performance (Spearman's ρ 0.53, $p < 0.001$). Actual ITE performance quintiles were contrasted with the resident's predicted ITE quintile relative to peers in Figure 2. The mean values of predicted ITE level were greatest in the middle of the distribution at 55.0% of the residents. The mean predicted relative quintile is most closely matched to actual performance quintile in the middle of the distribution as well (third quintile had a mean predicted quintile of 3.0). A total of 8.5% of residents in the first (lowest) quintile and 15.7% of residents in the fifth (highest) quintile correctly predicted their performance compared to their peers. Actual ITE performance quintiles are contrasted with the resident's predicted ITE quintile in Table 2. The highest percentage of correct predictions was in the middle of the distribution, meaning the third quintile of actual

TABLE 2 Confusion matrix (consistency table) of predicted and actual in-training examination (ITE) quintiles.

Actual ITE quintile	Predicted ITE quintile at level					Correct (%)
	1	2	3	4	5	
1	80	312	482	62	3	8.5
2	22	200	489	88	5	24.9
3	9	121	493	146	14	63.0
4	9	63	398	209	24	29.7
5	2	25	270	247	101	15.7

performance was able to correctly place themselves in the third quintile 63.0% of the time. The highest quintile underpredicted their performance while lowest quintile overpredicted their performance (Figure 2).

4 | LIMITATIONS

There are several limitations to this study. First, the cross-sectional study design is vulnerable to information and selection bias, specifically recall bias for residents' assessment of their relative performance and non-respondent bias where differential selection may occur through omission of the voluntary survey questions. We consider that the historical high proportion of response to the post-ITE survey was sufficient in that our sample is representative of the population of interest and adequately robust to assess for the presence of the DKE and to minimize the impact of information bias. Because this study only evaluates performance on the ITE, it does not evaluate how residents perceive their own clinical knowledge or performance or how they compare their clinical performance compared to peers. Evaluation of a resident's performance in and perception of medical knowledge compared to peers is one measure, and not the totality, of residency performance. Additionally, an individual resident taking the ITE is making a best guess on how they compare to peers in performance, and their perception might be impacted by proximity bias of the context of their own residency program and their limited knowledge of other programs or EM residents. However, our results are informative as a first step in addressing gaps between perceived and actual performance on the ABEM ITE among EM residents relative to their peers. Finally, it is important to note that this study operates under the assumption that the DKE can be quantified as a measurable cognitive bias and does not dismiss the evidence demonstrating that the DKE is a statistical artifact requiring further investigation.¹⁵

EM residents demonstrate accurate self-assessment of their performance on the ABEM ITE; however, the DKE is present when comparing their self-assessments to their peers. Notably, lower-performing EM residents tended to overestimate their performance, with the most significant DKE effect observed among the lowest-performing residents. Conversely, the highest-performing residents tended to underestimate their relative performance. These results can inform how the ABEM ITE might be used as a formative educational instrument. Investigation into the association of DKE with gender, race, and experience, as well as the impact of self-assessment mismatch on ABEM qualifying exam results will be the focus of future evaluation.

5 | DISCUSSION

The DKE can occur in any area of knowledge or skill, including academic, social, and emotional intelligence. It arises when people lack the metacognitive ability to accurately evaluate their own skills and knowledge. This is the first report evaluating DKE among EM residents taking the ABEM ITE. We found that EM residents did not exhibit the classic DKE when predicting their own absolute raw ITE. While all residents underpredicted their absolute performance, the highest scoring residents had the biggest gap between prediction and absolute score.

Residents exhibit more classic signs of DKE when estimating their performance relative to their peers (i.e., predicted percentile compared

to Post Graduate Year (PGY) peers). The lowest-performing residents overestimated their performance (to a larger degree) compared to peers, while the highest-performing residents underestimated their performance (to a smaller degree) compared to peers.

These findings align with others, showing limited or no DKE in absolute terms, but a more magnified effect in relative terms.¹⁶ Our data suggest that high-performing residents consistently underpredict their scores in both absolutely and relatively terms. Among the residents with the lowest scores, the DKE appears to be more pronounced only relative to peers, overestimating their abilities or underestimating their peers.

Understanding the DKE is crucial when preparing for high-stakes exams. Overconfidence can lead to under-preparation, while under-confidence might cause unnecessary stress and anxiety. Awareness of the DKE is essential for accurate self-assessment, effective study strategies, better utilization of feedback, and reduced test anxiety, all of which may contribute to better outcomes.

In clinical practice reflection and accurate self-assessment are pivotal for recognizing and addressing knowledge gaps.⁹ The DKE can interfere with the ability of those with the largest gaps to identify their deficits. This can lead to overconfidence, poor decision making, failure to seek assistance with patient management, and the failure to recognize the need for further learning and development. Informing EM residents of their tendency to overestimate or to underestimate their abilities in relation to their peers could provide insight into their self-assessment and help them to optimize knowledge acquisition and self-awareness in professional development. It could also encourage improved self-directed learning to close knowledge gaps.

Assessment is a key component of the clinical learning environment.¹⁷ Mitigating the DKE involves implementing strategies that promote self-awareness, accurate self-assessment, and continuous learning. Healthy habits of life-long learning thrive in environments where a combination of self-assessments, workplace-based assessments,¹⁸ such as daily evaluation cards, entrustable professional activities,¹⁹ and independent assessments, such as the ABEM ITE, occur with minimal distortion between assessment and performance.

Recognizing the DKE allows individuals to gain insight into their own limitations and strengths. This awareness fosters personal development by encouraging individuals to acknowledge areas where they may lack expertise and seek opportunities for improvement.⁴ By understanding the potential for overestimating their abilities, individuals can also approach decision making with greater caution and humility, seeking input from others and avoiding the pitfalls of overconfidence.²⁰

Discussions regarding humility and confidence in the clinical learning environment are increasing.^{9,21,22} Intellectual humility (the willingness to recognize the limits of one's knowledge and the openness to learning from others) along with self-confidence can play significant roles in shaping a resident's learning and professional development.²² In addition, further studies of the relationship between DKE and intellectual humility may be helpful in gaining insight into how residents approach their learning, patient care, and interactions with colleagues.

AUTHOR CONTRIBUTIONS

Study design, data analysis/interpretation, manuscript development, and review: Felix Ankel, Theodore Gaeta and Earl Reisdorff. *Manuscript review:* Melissa Barton, Kim M. Feldhaus, Marianne Gausche-Hill, and Deepi Goyal. *Study design, data analysis, and manuscript review:* Kevin Joldersma. *Manuscript development and manuscript review:* Chadd K. Kraus.

ACKNOWLEDGMENTS

The authors wish to thank Ms. Lyndsay Tyler for her valued work coordinating the development of the manuscript and Ms. Frances Spring for her valued assistance in preparing and submitting the manuscript.

CONFLICT OF INTEREST STATEMENT

T.J.G., K.M.F., M.G.H., D.G., and F.A. serve on the American Board of Emergency Medicine Board of Directors. E.J.R., M.B., K.B.J., and C.K.K. are employed by the American Board of Emergency Medicine. The American Board of Emergency Medicine receives revenue from administering the ITE used in this study.

REFERENCES

- Wolff M, Santen SA, Hopson LR, Hemphill RR, Farrell SE. What's the evidence: self-assessment implications for life-long learning in emergency medicine. *J Emerg Med*. 2017;53(1):116-120.
- Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;296(9):1094-1096.
- Sadosty AT, Bellolio MF, Laack TA, Luke A, Weaver A, Goyal DG. Simulation-based emergency medicine resident self-assessment. *J Emerg Med*. 2011;41(6):679-685.
- 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.
- Burson KA, Larrick RP, Klayman J. Skilled or unskilled, but still unaware of it: how perceptions of difficulty drive miscalibration in relative comparisons. *J Pers Soc Psychol*. 2006;90(1):60-77.
- 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.
- Osterhage JL. Persistent miscalibration for low and high achievers despite practice test feedback in an introductory biology course. *J Microbiol Biol Educ*. 2021;22(2):1-11.
- Hartigan S, Brooks M, Hartley S, Miller RE, Santen SA, Hemphill RR. Review of the basics of cognitive error in emergency medicine: still no easy answers. *West J Emerg Med*. 2020;21(6):125-131.
- Rahmani M. Medical trainees and the Dunning-Kruger effect: when they don't know what they don't know. *J Grad Med Edu*. 2020;12(5):532-534.
- TenEyck L. Dunning-Kruger effect. In: Raz M, Pouryahya P, eds. *Decision Making in Emergency Medicine: Biases, Errors and Solutions*. Singapore: Springer Nature; 2021:123-128.
- McIntosh RD, Fowler EA, Lyu T, Della Sala S. Wise up: clarifying the role of metacognition in the Dunning-Kruger effect. *J Exp Psychol Gen*. 2019;148(11):1882-1897.

- Dunning D. Chapter 5: the Dunning-Kruger effect: on being ignorant of one's own ignorance. *Adv Exp Soc Psychol*. 2011;44:247-296.
- StataCorp. *Stata Statistical Software: Release 18*. 2023. College Station, TX: StataCorp LLC.
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. 2024. <https://www.R-project.org/>
- Magnus JR, Peresetsky AA. A statistical explanation of the Dunning-Kruger effect. *Front Psychol*. 2022;13:840180.
- Mazor M, Fleming SM. The Dunning-Kruger effect revisited. *Nat Hum Behav*. 2021;5(6):677-678.
- Schuwirth LWT, van der Vleuten CPM. A history of assessment in medical education. *Adv Health Sci Educ Theory Pract*. 2020;25(5):1045-1056.
- Bandiera G, Hall AK. Capturing the forest and the trees: workplace-based assessment tools in emergency medicine. *CJEM*. 2021;23(3):265-266.
- Ten Cate O, Hart D, Ankel F, et al. Entrustment decision making in clinical training. *Acad Med*. 2016;91(2):191-198.
- Burson Katherine A, Larrick RP, Klayman J. Skilled or unskilled, but still unaware of it: how perceptions of difficulty drive miscalibration in relative comparisons. *J Pers Soc Psychol*. 2006;90(1):60-77.
- Hong LT, Gavaza P, Koch J, de la Pena I. An explorative study on intellectual humility differences among health professional students: implications for interprofessional education and collaboration. *J Interprofessional Educ Prac*. 2023;33:100674. doi:10.1016/j.xjep.2023.100674
- Krumrei-Mancuso EJ, Haggard MC, LaBouff JP, Rowatt WC. Links between intellectual humility and acquiring knowledge. *J Positive Psychol*. 2020;15(2):155-170.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Gaeta TJ, Reisdorff E, Barton M, et al. The Dunning-Kruger effect in resident predicted and actual performance on the American Board of Emergency Medicine in-training examination. *JACEP Open*. 2024;5:e13305. <https://doi.org/10.1002/emp2.13305>

AUTHOR BIOGRAPHY



Theodore J. Gaeta, DO, MPH, is the Vice-chairman and Program Director in Emergency Medicine and Chief Research Officer at New York-Presbyterian Brooklyn Methodist Hospital in Brooklyn, New York, USA. He is an Associate Professor of Clinical Emergency Medicine at Weill

Cornell Medicine.