

# Self-estimation of surgical skills and competencies based on the learning curve theory in medical residents and fellows

## Autoestimación de habilidades y competencias quirúrgicas basadas en la teoría de la curva de aprendizaje en médicos residentes y becarios

Raul A. Borracci<sup>1,2\*</sup>, José M. Alvarez-Gallesio<sup>2</sup>, Graciana Ciambone<sup>2</sup> and Norberto A. Mezzadri<sup>3</sup>

<sup>1</sup>Biostatistics, School of Medicine, Austral University, Buenos Aires, Argentina; <sup>2</sup>Medical Education Research Laboratory, Deutsches Hospital, Buenos Aires, Argentina; <sup>3</sup>Department of Surgery, Deutsches Hospital, Buenos Aires, Argentina

### Abstract

**Objective:** The aim was to explore how in-training junior physicians perceive their surgical performance compared with the one externally rated by their senior surgeon trainers, using a general learning curve model. **Methods:** Between April and June 2018, a prospective study was conducted at a community hospital associated with a school of medicine. To assess how in-training physicians estimated their surgical performance, 48 surgical residents and fellows were invited to choose one among six options using a scale ranging from "novice" to "automatic expert." In addition, five senior surgeons who supervised the residents/fellows were asked to give their own opinions on each surveyed physician's expertise level, according to the same categories. Concordance analysis was done to compare residents' and fellows' self-perceived skills and their actual performance as estimated by senior surgeons. **Results:** Self-assessments tended to overestimate residents' and fellows' position on the learning curve; particularly for "proficient" over "competent," and for "automatic expert" over "expert" categories ( $p = 0.025$ ). The average degree of agreement among senior physicians was 50.0%. Comparison between residents' and fellows' perceived skills and their performances as estimated by senior surgeons showed a weak concordance ( $\kappa = 0.494$ , 95% confidence interval 0.359-0.631,  $p < 0.0001$ ). **Conclusions:** Nearly 51% of the residents/fellows included in some surgical specialty training program overestimated his/her actual performance as evaluated by classical learning curve categories. Underestimation of self-assessed performance was also observed in 17% of respondents. A better feedback from expert observers to in-training surgeons could result in a more accurate self-perception of their real surgical skills and competencies.

**KEY WORDS:** Self-assessment. Expert-assessment. Technical skills. Learning curve. Surgery.

### Resumen

**Objetivo:** Evaluar cómo los médicos en formación (juniors) perciben su propio desempeño quirúrgico en comparación con la calificación otorgada por sus instructores (seniors) según un modelo de curva de aprendizaje. **Métodos:** Entre abril y junio de 2018 se realizó un estudio prospectivo en un hospital comunitario. Para evaluar cómo los médicos juniors estimaban su propio desempeño, 48 residentes/becarios de especialidades quirúrgicas eligieron una entre seis opciones excluyentes en una escala entre «novicio» y «experto automático». Además, cinco cirujanos que supervisaban a los residentes/becarios dieron sus propias opiniones sobre el nivel de desempeño de cada médico encuestado, usando las mismas categorías. Se realizó un análisis de concordancia para comparar las habilidades autopercebidas y el desempeño real según lo estimado por los cirujanos seniors. **Resultados:** Cuarenta y siete juniors y 50 seniors completaron la encuesta. El 51% sobrestimó y el 17% subestimó su ubicación en la curva de aprendizaje con respecto a los observadores externos ( $p = 0.025$ ). El grado promedio de acuerdo entre seniors fue del 50%. La comparación entre la autopercepción de los juniors con respecto a sus observadores seniors mostró una con-

### Correspondence:

\*Raul A. Borracci  
La Pampa 3030  
1428 Buenos Aires, Argentina  
E-mail: raborracci@gmail.com

Date of reception: 15-08-2018  
Date of acceptance: 10-10-2018  
DOI:10.24875/CIRU.18000664

Cir Cir. 2019;87:416-422  
Contents available at PubMed  
www.cirugiyacirujanos.com

cordancia pobre ( $kappa = 0.494$ ; intervalo de confianza del 95% [IC 95%]: 0.359-0.631;  $p < 0.0001$ ; sesgo promedio de Bland-Altman: 0.40; IC 95%: 0.11-0.70). **Conclusiones:** La mitad de los residentes/fellows sobrestimó, y uno de cada seis subestimó, su verdadera ubicación en la curva de aprendizaje en comparación a la opinión de los seniors. Un mejor conocimiento de la existencia de este sesgo de estimación del propio desempeño podría redundar en una mejor confiabilidad del juicio médico.

**PALABRAS CLAVE:** Autoevaluación. Cirugía. Curva de aprendizaje. Evaluación por experto. Habilidades técnicas.

## Introduction

The safe surgical practice relies on surgeons being aware of their own skill sets and capabilities, as well as on the acknowledgment of their limitations. Particularly among in-training physicians, accurate self-assessment of both confidence and competence is important goals to ensure an adequate learning process and safer patient health care<sup>1</sup>. Nevertheless, the ubiquitous over/under-confidence bias may impose a miscalibration between self-judgments and the objective accuracy of those judgments. Thus, Pallier et al<sup>2</sup>. have identified overconfidence in three different ways: (1) as an overestimation of one's actual performance, (2) as an over placement of one's performance relative to others, and (3) as an excessive certainty regarding the accuracy of one's beliefs or knowledge, known as overprecision. Studies on physicians found that their self-assessment of clinical skills did not correlate well with an external evaluation of the same competencies, and the most inaccurate self-assessments were observed in the physicians who expressed the highest confidence level or those who were externally-rated to be the lowest<sup>3</sup>. On the other hand, studies conducted on junior doctors showed more variable results in terms of the correlation between self-perceived and objectively measured or observed competency, with poorer correlations in practical clinical skills<sup>4-6</sup>. Although weak or no associations between physicians' self-rated and external assessments have been often observed<sup>3</sup>, evaluating self-perceived competence may provide an indication on the subject's motivation in maintaining and improving the skills concerned, and furthermore, it is considered an important component of self-efficacy<sup>7</sup>. Several classical assessment tools and methods have been used in studies reporting self-assessment of surgical skills and competencies: operative component rating scale<sup>8</sup>, global rating scale<sup>9</sup>, global score<sup>10</sup>, visual analogue scale<sup>11</sup>, standardized forms<sup>12</sup>, blinded or not blinded direct observation<sup>13,14</sup>, single or multiple observers<sup>15</sup>, video playback analysis<sup>16</sup>, objective structured assessment of surgical skills<sup>14,17</sup>, hierarchical task analysis<sup>10</sup>, self-assessment score of performance<sup>15</sup>, competency

assessment tool<sup>18</sup>, bench models<sup>11,13</sup>, virtual reality simulators<sup>13,19</sup>, live animal models<sup>9</sup>, and live operating setting<sup>10</sup>. All these different approaches have shown that the evidence of self-estimated accuracy of surgical technical skills is still contradictory<sup>20,21</sup>.

Meanwhile, the learning curve theory has been recently revisited and promoted as a valuable method to assess medical competencies<sup>22</sup>. Learning curve models are useful to assess an individual physician's progress toward his/her medical capabilities in patient care, by graphically representing the relationship between the learning effort and the resultant learning outcomes. Alternatively, surgical learning curve describes the relationship between deliberate practice and subsequent performance through a classical S-shaped curve that divides a series of ascending categories of expertise. Evidence supporting the validity of a learning curve as a useful tool to assess skill acquisition basically relies on the Dreyfus<sup>23</sup> and Ericsson<sup>24,25</sup> models which describe expertise development as a progression through several stages, from a novice who is not allowed to practice on patients to a reflective expert who functions at the highest levels.

Based on this theoretical framework, we hypothesized that in-training residents and fellows could over or underestimate their actual surgical skills compared with their performances perceived by an external expert observer. Therefore, the aim of this study was to explore how in-training junior physicians perceive their surgical performance compared with the one externally-rated by their senior surgeon trainers, using a general learning curve model.

## Material and methods

Between April and June 2018, a prospective study was conducted at a community hospital associated with the Buenos Aires University School of Medicine. To assess how in-training young physicians estimated their surgical performance, 48 first- to fourth-year surgical residents and fellows were invited to choose one among six exclusive options, which were intended to summarize their own perceived performance or skills

as in-training surgeons at the moment of the survey. Residents and fellows were asked to place themselves in one of the following learning curve categories:

- Novice: *I have no skill or experience to perform any surgical procedure.*
- Advance: *I can practice some surgical procedures with full supervision.*
- Competent: *I can practice some surgical procedures with supervision on call.*
- Proficient: *I can practice some surgical procedures without supervision.*
- Expert: *I can supervise others to practice some surgical procedures.*
- Automatic expert: *I can practice some surgical procedures automatically.*

To minimize biased selection, residents had only access to definitions, but not to the names of categories. These learning curve theoretical approaches and definitions were adopted from Pusic et al<sup>22</sup>. To avoid inconsistent opinions of junior physicians with null surgical experience, 1<sup>st</sup>-year residents participated in the survey when they had completed at least 10 months within the surgical residency program. After selecting their own perceived surgical performance, five selected senior surgeons (multiple observers design) who supervised the residents and fellows, were asked to give their own opinions about the expertise level reached by each surveyed in-training physician, according to the same learning curve categories. Opinions were considered to be expressed in a double-blind way since neither residents/fellows nor surgeons knew the existence of a cross-evaluation. In this case, statistical analysis was done by comparing the level of concordance between residents' and fellows' own perceived skills and their actual performances as estimated by senior assistant surgeons. From a traditional viewpoint, perceived skills were defined as the self-reported confidence level, and estimated performance as the observed competence<sup>17</sup>.

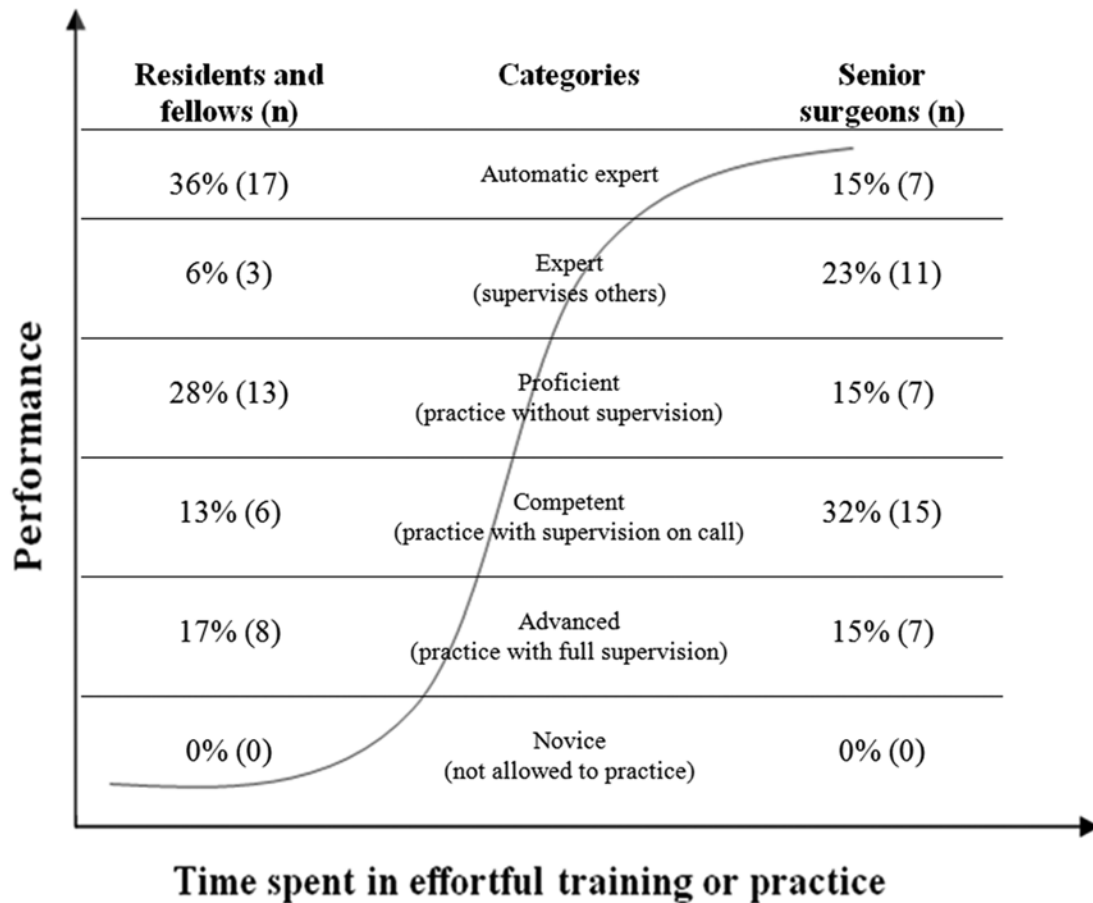
Participants were assured confidentiality in responding to the questionnaire. All respondents voluntarily participated in the study after being explained its purpose and expressed consent by filling the form. All personal identifiers were removed or disguised so the physicians described were not identifiable and could not be identified through the details of the study. Heads of the medical training institution provided access to the residents and fellows after ethical approval of the protocol. Ethical clearance for this study was granted by the Institutional Review Board of the Deutsches Hospital of Buenos Aires.

## Statistical analysis

Cohen's kappa statistic and weighted kappa with Cicchetti's weighting scheme were used to assess concordance between residents' and fellows perceived skills and their performances as estimated by senior surgeons. The median value of multiple external observers was used for the purpose of analysis. Qualitative interpretation of kappa indexes was based on current recommendations<sup>26</sup>. 95% confidence intervals (95% CI) for concordance indexes were also calculated. Since the differences between the learning curve ordinal categories are critical for surgical skill development, we preferred to weigh kappa statistic with Cicchetti's proportional weighting instead of exponential quadratic weighting. The sample size for weighted kappa analysis was estimated with  $n = 2c^2$ , where  $c$  is the number of categories<sup>27</sup>. Since the first category (novice) was expected not to be selected by in-training physicians, then the sample size calculated was  $n = 50$  individuals. The degree of agreement (inter-rater reliability, [IRR]) among multiple external observers (senior surgeons) regarding junior doctors' performance level was expressed as percent agreement and intraclass correlation coefficient. Percent agreement was calculated as the number of agreement scores divided by the total number of scores. Overall comparison between junior and senior physician responses was done with Yates' Chi-square test for 4 degrees of freedom, according to the number of selected categories. Continuous variables were expressed as mean and standard deviation (SD). Statistical analysis was performed with EPIDAT, Version 4.1 (Xunta de Galicia-PAHO/WHO), and SPSS Statistics for Windows, Version 17.0 (Chicago: SPSS Inc.) and a two-tailed  $p \leq 0.05$  was considered statistically significant.

## Results

Forty-seven out of 48 first- to fourth-year surgical residents and fellows (98%), and 50 senior surgeons (5 for each specialty) completely responded the survey. The study included the following surgical specialties and number of participants: general ( $n = 11$ ), colorectal ( $n = 2$ ), liver ( $n = 1$ ), plastic ( $n = 2$ ), cardiovascular ( $n = 2$ ), neurological ( $n = 2$ ), urological ( $n = 5$ ), gynecological and obstetrics ( $n = 7$ ), orthopedic ( $n = 10$ ), and ophthalmological ( $n = 5$ ) surgery. Mean age of residents and fellows was 29.6 years (SD 2.9), and 30 (64%) were male.



**Figure 1.** Residents' and fellows' responses to self-estimation of their surgical skills and competencies (left column), compared with their actual performances as estimated by senior assistant surgeons (right column) based on learning curve categories (Yates' Chi-square = 11.2, degrees of freedom = 4,  $p = 0.025$ ).

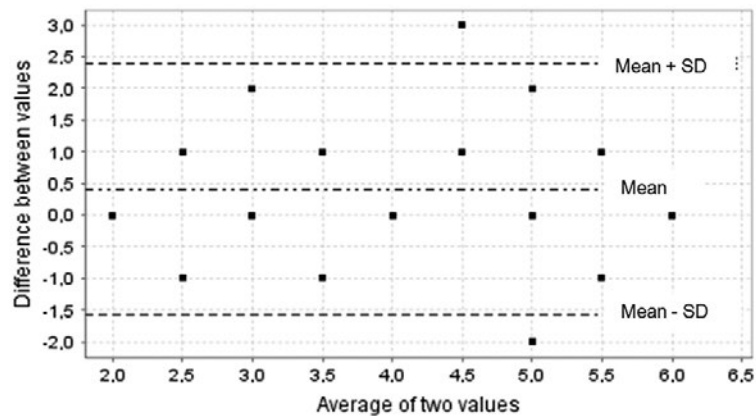
Figure 1 shows the response of residents and fellows to self-estimation of their surgical skills and competencies compared with their actual performances as estimated by senior assistant surgeons, based on learning curve categories. Globally, self-assessments tended to overestimate their positions on the learning curve; particularly for "proficient" over "competent," and for "automatic expert" over "expert" categories ( $p = 0.025$ ). 24 (51%) and 8 (17%) residents and fellows overestimated and underestimated his/her performance, respectively. Overestimation rate was 38% (10/26) for first- to third-year residents versus 64% (14/22) for the rest of respondents ( $p = 0.148$ ), whereas underestimation was 19% ( $n = 5$ ) versus 14% ( $n = 3$ ), respectively ( $p = 0.897$ ). Average degree of agreement among senior physician responses was 50.0% (95% CI 43.7-56.3%) (intraclass correlation coefficient = 0.737, 95% CI 0.637-0.825). Comparison between residents' and fellows' perceived skills and their performances as estimated by senior surgeons

showed a poor to weak concordance according to kappa measures (kappa = 0.174, 95% CI 0.019-0.328,  $p = 0.007$  and weighted kappa = 0.494, 95% CI 0.359-0.631,  $p < 0.0001$ ). The Bland-Altman plot of the difference between self-estimation and external evaluation of surgical skills is shown in figure 2. Average bias between paired values was 0.40 (95% CI 0.11-0.70), demonstrating lack of concordance between in-training doctors' and senior surgeons' opinions. The positive deviation of the difference between responses revealed a global overestimation of their surgical skills and competencies as seen by junior physicians.

## Discussion

A double-blind cross-validation design study with multiple external observers was conducted to assess self-estimation of surgical competencies among in-training junior surgeons, compared with the external evaluation made by their senior surgeon trainers.





**Figure 2.** Bland-Altman plot of the difference between self-estimation and external estimation of surgical skills and competencies of in-training junior physicians. Average bias between paired values was 0.40 (95% confidence interval [CI] 0.11-0.70). Since this CI does not include the zero value, it can, therefore, be considered that there is no significant concordance between both evaluations. Most pairs of values overlap in the plot; therefore, the number of points on the graph seems to be less than the total sample size.

Concordance analysis demonstrated that residents and fellows of surgical specialties tended, in general, to overestimate their current performances regarding learning curve categories, when compared with external observers' opinions. The perceived level of competence among junior doctors revealed a high proportion of them thinking they are able "to practice without supervision," or being an "automatic expert" on the top stratum of the learning curve when performing some surgical procedures.

In medicine, the development of expertise requires the recognition of one's capabilities and limitations<sup>8</sup>. Safe clinical practice depends on being able to recognize the limits of one's competence so that the doctor does not only take unnecessary risks but also underconfidence would make physicians unable to act to prevent critical incidents<sup>1</sup>. Therefore, from a patient's safety perspective, the relationship between confidence and competence is crucial.

It is controversial whether self-assessment is an accurate form of technical skill appraisal in surgical specialties<sup>20</sup>. In general surgery, four studies reported that candidates' self-assessment and expert independent evaluation correlate poorly, with trainees overestimating their abilities<sup>9,11,13,16</sup>. For example, surgeons consistently overestimated their performance during a laparoscopic colectomy course as measured by a reliable global rating scale<sup>9</sup>. Using a 5-point scale from "novice" to "expert," Morgan and Cleave-Hogg<sup>28</sup> found that the level of confidence of medical students had no predictive value in performance assessment on anesthesia simulated scenarios. A discrepancy was also observed among urology

residents' perceptions of their skills' proficiency, compared with faculty members' evaluations<sup>29</sup>. An identical lack of concordance was reported when assessing operative skills of pediatric neurosurgery residents<sup>30</sup>. Conversely, another nine investigations reported good self-assessment accuracy in general surgery<sup>8,10,12,14,15,18,19,31,32</sup>, and in a pilot study, orthopedic surgery residents could successfully self-assess their performance using a milestones-based method<sup>33</sup>. Most of these studies included only one to three external observers, with IRR fluctuating between 0.61 and 1. In the current study, a lower IRR was expected since the opinion of five external observers was included for each junior doctor.

People tend to overestimate their ability in many different domains, with this overestimation increasing with harder tasks and decreasing with easier tasks<sup>34,35</sup>. Some evidence suggests that self-appraisal is more accurate with increased experience<sup>16,31</sup>, surgical training level and age<sup>14</sup>. Conversely, there is other evidence regarding an increase of underconfidence with practice; this counterintuitive effect seems to depend on the awareness of self-limitation in task performance<sup>36,37</sup>. Although we have observed a global underestimation rate of 17%, we did not find that paradoxical effect when comparing postgraduate one- to third-year versus the rest of the participants.

Some authors suggested that self-assessment of a cognitive task may be fundamentally different from an objective technical task. This is based on the notion that the performance of technical labor, unlike a cognitive one, can be judged through immediate or direct

feedback provided by the outcome of that labor<sup>21,38</sup>. Thus, the agreement between self and external assessment for cognitive tasks may be different for technical tasks. Hence, strategies to improve the agreement between self and external assessment in the context of surgical training should include high-quality, timely, coherent, and non-threatening external feedback from expert observers to trainees<sup>38</sup>.

This study has some limitations. First, the possible ambiguity of some statements of the questionnaire would be offset by the simultaneous application of the same survey to junior and senior physicians. Probably, junior doctors' estimation level would vary if they confronted a real surgical situation than a paper-based survey. Another limitation is that the demonstration of misplaced estimation among residents and fellows does not necessarily mean that consequences or benefits are derived from it, or that these biases are necessarily a problem. Some bias could emerge from the assignment of certain senior surgeons as observers capable of judging the participants, and probably greater standardization should be required for the external assessors. Although it is unlikely that one standard self-assessment tool can be suitably applied to all technical procedures, we used a global approach based on learning curve categories to achieve a general image of the perceived estimation of in-training junior physicians' surgical performance. Although confidence and competence are linearly associated, there is a critical difference in whether trainees have gained a greater belief in their abilities at carrying out a particular skill and whether they are technically more proficient in putting them into practice<sup>1</sup>.

## Conclusions

Comparison of self-reported estimation of residents' and fellows' surgical skills with the observed competence estimated by their senior surgeon trainers showed poor concordance. About half of the residents and fellows included in some surgical specialty training program overestimated his/her actual performance as assessed by classical learning curve categories. Nevertheless, underestimation of self-assessed performance was also observed in almost one-fifth of the respondents. An increased awareness of the existence of over- and underestimation effects can increase the reliability of medical judgment. An improved feedback from expert observers to in-training surgeons could result in a more accurate self-perception of their real surgical skills and competencies.

## Funding

The authors have not received financial support for this work.

## Conflicts of interest

The authors declare that they have no conflicts of interest.

## Ethical disclosures

**Protection of human and animal subjects.** The authors declare that the procedures followed were in accordance with the regulations of the relevant clinical research ethics committee and with those of the Code of Ethics of the World Medical Association (Declaration of Helsinki).

**Confidentiality of data.** The authors declare that they have followed the protocols of their work center on the publication of patient data.

**Right to privacy and informed consent.** The authors have obtained the written informed consent of the patients or subjects mentioned in the article. The corresponding author is in possession of this document.

## References

1. Roland D, Matheson D, Coats T, Martin G. A qualitative study of self-evaluation of junior doctor performance: is perceived safeness a more useful metric than confidence and competence? *BMJ Open*. 2015;5:e008521.
2. Pallier G, Wilkinson R, Danthiir V, Kleitman S, Knezevic G, Stankov L, et al. The role of individual differences in the accuracy of confidence judgments. *J Gen Psychol*. 2002;129:257-99.
3. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L, et al. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;296:1094-102.
4. Barnsley L, Lyon PM, Ralston SJ, Hibbert EJ, Cunningham I, Gordon FC, et al. Clinical skills in junior medical officers: a comparison of self-reported confidence and observed competence. *Med Educ*. 2004;38:358-67.
5. Coberly L, Goldenhar LM. Ready or not, here they come: acting interns experience and perceived competency performing basic medical procedures. *J Gen Intern Med*. 2007;22:491-4.
6. Jones A, McArdle PJ, O'Neill PA. How well prepared are graduates for the role of pre-registration house officer? A comparison of the perceptions of new graduates and educational supervisors. *Med Educ*. 2001;35:578-84.
7. 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.
8. Ward M, MacRae H, Schlachta C, Mamazza J, Poulin E, Reznick R, et al. Resident self-assessment of operative performance. *Am J Surg*. 2003;185:521-4.
9. Sidhu RS, Vikis E, Cheifetz R, Phang T. Self-assessment during a 2-day laparoscopic colectomy course: can surgeons judge how well they are learning new skills? *Am J Surg*. 2006;191:677-81.
10. Sarker SK, Hutchinson R, Chang A, Vincent C, Darzi AW. Self-appraisal hierarchical task analysis of laparoscopic surgery performed by expert surgeons. *Surg Endosc*. 2006;20:636-40.
11. van Empel PJ, Verdam MG, Huirne JA, Bonjer HJ, Meijerink WJ, Scheele F, et al. Open knot-tying skills: resident skills assessed. *J Obstet Gynaecol Res*. 2013;39:1030-6.
12. Brewster LP, Risucci DA, Joehi RJ, Littooy FN, Temeck BK, Blair PG, et al. Comparison of resident self-assessments with trained faculty and

- standardized patient assessments of clinical and technical skills in a structured educational module. *Am J Surg.* 2008;195:1-4.
13. Pandey VA, Wolfe JH, Black SA, Cairls M, Liapis CD, Bergqvist D, et al. Self-assessment of technical skill in surgery: the need for expert feedback. *Ann R Coll Surg Engl.* 2008;90:286-90.
14. de Blacam C, O'Keefe DA, Nugent E, Doherty E, Traynor O. Are residents accurate in their assessments of their own surgical skills? *Am J Surg.* 2012;204:724-31.
15. Tedesco MM, Pak JJ, Harris EJ Jr, Krummel TM, Dalman RL, Lee JT, et al. Simulation-based endovascular skills assessment: the future of credentialing? *J Vasc Surg.* 2008;47:1008-1.
16. Hu Y, Tiemann D, Michael Brunt L. Video self-assessment of basic suturing and knot tying skills by novice trainees. *J Surg Educ.* 2013;70:279-83.
17. Quick JA, Kudav V, Doty J, Crane M, Bukoski AD, Bennett BJ, et al. Surgical resident technical skill self-evaluation: increased precision with training progression. *J Surg Res.* 2017;218:144-9.
18. Ganni S, Chmarra MK, Goossens RH, Jakimowicz JJ. Self-assessment in laparoscopic surgical skills training: is it reliable? *Surg Endosc.* 2017;31:2451-6.
19. Arora S, Miskovic D, Hull L, Moorthy K, Aggarwal R, Johannsson H, et al. Self vs expert assessment of technical and non-technical skills in high fidelity simulation. *Am J Surg.* 2011;202:500-6.
20. Rizan C, Ansell J, Tilston TW, Warren N, Torkington J. Are general surgeons able to accurately self-assess their level of technical skills? *Ann R Coll Surg Engl.* 2015;97:549-55.
21. Zevin B. Self versus external assessment for technical tasks in surgery: a narrative review. *J Grad Med Educ.* 2012;4:417-24.
22. Pusic MV, Boutis K, Hatala R, Cook DA. Learning curves in health professions education. *Acad Med.* 2015;90:1034-42.
23. Dreyfus SE, Dreyfus HK. A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition. Berkeley, California: University of California, Berkeley; 1980.
24. Ericsson K. The acquisition of expert performance as problem solving. In: Davidson J, Sternberg R, editors. *The Psychology of Problem Solving.* 1st ed. Cambridge, England: Cambridge University Press; 2003. p. 44-83.
25. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med.* 2008;15:988-94.
26. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology: a Basic Science for Clinical Medicine.* 2nd ed. Boston: Brown and Co.; 1991.
27. Cicchetti DV. Testing the normal approximation and minimal sample size requirements of weighted kappa when the number of categories is large. *Appl Psychol Meas.* 1981;5:101-4.
28. Morgan PJ, Cleave-Hogg D. Comparison between medical students experience, confidence and competence. *Med Educ.* 2002;36:534-9.
29. Mitchell RE, Clark PE, Scarpero HM. Assessing the surgical skills of urology residents after preurology general surgery training: the surgical skills learning needs of new urology residents. *J Surg Educ.* 2011;68:341-6.
30. Aldave G, Hansen D, Briceño V, Luerksen TG, Jea A. Assessing residents' operative skills for external ventricular drain placement and shunt surgery in pediatric neurosurgery. *J Neurosurg Pediatr.* 2017;19:377-83.
31. Moorthy K, Munz Y, Adams S, Pandey V, Darzi A, Imperial College St. Mary's Hospital Simulation Group. et al. Self-assessment of performance among surgical trainees during simulated procedures in a simulated operating theater. *Am J Surg.* 2006;192:114-8.
32. Munz Y, Moorthy K, Bann S, Shah J, Ivanova S, Darzi SA, et al. Ceiling effect in technical skills of surgical residents. *Am J Surg.* 2004;188:294-300.
33. Bradley KE, Andolsek KM. A pilot study of orthopaedic resident self-assessment using a milestones survey just prior to milestones implementation. *Int J Med Educ.* 2016;7:11-8.
34. Meyer AN, Payne VL, Meeks DW, Rao R, Singh H. Physicians diagnostic accuracy, confidence, and resource requests: a vignette study. *JAMA Intern Med.* 2013;173:1952-8.
35. 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;79:216-47.
36. 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;131:147-62.
37. Finn B, Metcalfe J. The role of memory for past test in the underconfidence with practice effect. *J Exp Psychol Learn Mem Cogn.* 2007;33:238-44.
38. Eva KW, Regehr G. Exploring the divergence between self-assessment and self-monitoring. *Adv Health Sci Educ Theory Pract.* 2011;16:311-29.