

BMJ Open How well do doctors think they perform on the General Medical Council's Tests of Competence pilot examinations? A cross-sectional study

Leila Mehdizadeh,¹ Alison Sturrock,¹ Gil Myers,¹ Yasmin Khatib,² Jane Dacre³

To cite: Mehdizadeh L, Sturrock A, Myers G, *et al.* How well do doctors think they perform on the General Medical Council's Tests of Competence pilot examinations? A cross-sectional study. *BMJ Open* 2014;**4**:e004131. doi:10.1136/bmjopen-2013-004131

► Prepublication history and additional material for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2013-004131>).

Received 23 October 2013
Revised 17 December 2013
Accepted 19 December 2013



CrossMark

¹Division of UCL Medical School, University College London, London, UK

²Barts and the London School of Medicine and Dentistry, Centre for Psychiatry, Queen Mary University of London, London, UK

³Division of UCL Medical School, University College London, London, UK

Correspondence to

Dr Leila Mehdizadeh; l.mehdizadeh@ucl.ac.uk

ABSTRACT

Objective: To investigate how accurately doctors estimated their performance on the General Medical Council's Tests of Competence pilot examinations.

Design: A cross-sectional survey design using a questionnaire method.

Setting: University College London Medical School.

Participants: 524 medical doctors working in a range of clinical specialties between foundation year two and consultant level.

Main outcome measures: Estimated and actual total scores on a knowledge test and Observed Structured Clinical Examination (OSCE).

Results: The pattern of results for OSCE performance differed from the results for knowledge test performance. The majority of doctors significantly underestimated their OSCE performance. Whereas estimated knowledge test performance differed between high and low performers. Those who did particularly well significantly underestimated their knowledge test performance ($t(196)=-7.70$, $p<0.01$) and those who did less well significantly overestimated ($t(172)=6.09$, $p<0.01$). There were also significant differences between estimated and/or actual performance by gender, ethnicity and region of Primary Medical Qualification.

Conclusions: Doctors were more accurate in predicating their knowledge test performance than their OSCE performance. The association between estimated and actual knowledge test performance supports the established differences between high and low performers described in the behavioural sciences literature. This was not the case for the OSCE. The implications of the results to the revalidation process are discussed.

INTRODUCTION

The revalidation process of all UK doctors who hold a license to practise is now underway. Introduced by the General Medical Council (GMC), doctors must provide evidence at their annual appraisal that they continue to meet the standards set out in Good

Strengths and limitations of this study

- This is one of the first studies to look at how well a group of doctors think they perform on a set of examinations that have potentially significant real-world consequences.
- The large sample means that it has greater power than previous similar studies.
- The results have important implications to the revalidation process of doctors working in the UK.
- The majority of doctors performed well on both examinations, therefore patterns between high and low performers should be interpreted with caution.
- Results are not necessarily applicable to the wider medical community in the UK.

Medical Practice.^{1 2} This requires the doctor to reflect on their own knowledge and practise to be able to demonstrate their strengths and areas that need further development. While medical education in the UK aims to produce doctors who are reflective practitioners,^{2 3} evidence from the behavioural sciences has demonstrated the deficits in the ability to assess one's own competencies.⁴⁻¹⁰ Previous studies showed that psychology students who performed lowest on intellectual and social tasks displayed the least insight by overestimating their own performance. In contrast, the highest performers underestimated their performance.⁷ This pattern has been replicated in the medical context.

General practitioners were unable to accurately assess their knowledge of 20 typical clinical conditions.¹¹ Family medicine residents who performed best on a breaking bad news scenario were more likely to make accurate self-estimates than the lowest performers.¹² A systematic review of studies on the accuracy of doctors' self-assessment compared with objective measures of competence showed

that doctors who performed least well also self-assessed least well. These studies indicate that people in general, and medical doctors specifically, have a limited ability to accurately assess their own competence.¹³ In clinical practice, the idea that a poorly performing doctor could also lack awareness of their problems is concerning. Such a doctor may pose a risk to patient safety and may not engage in appropriate professional development activities. With the introduction of revalidation, the number of doctors referred to the GMC for Fitness to Practise (FtP) investigation may change. Revalidation has a strong component of self-evaluation and enforced reflection will offer doctors more opportunities to consider their performance and ways to remedy any weak areas.

There is a debate in the self-assessment literature concerning use of terminology and outcome measures. One perspective that has gained more recent attention is the need to distinguish the self-assessment approach from the self-monitoring approach to the investigation of self-perceived competence.^{14–17} The self-assessment approach investigates how well individuals can judge their personal competence against an objective measure of competence.¹⁶ Alternatively, the self-monitoring approach is interested in the extent that people show awareness of the limits in their competence during a given situation, and this can be measured according to an individual's behaviour, for example, taking longer time to think about a question of which they are unsure.¹⁶ It is important to clarify that this study looks at doctors' self-assessment ability after completing a set of examinations, but not on how well they can monitor their performance during the examinations. Therefore, this study does not include any outcomes of doctors' behaviour.

We measured the accuracy of doctors' self-predicted performance on the GMC's Tests of Competence (ToC) pilot examinations. ToC are used by the GMC to assess poorly performing doctors under FtP investigation. Before implementation, test content is piloted on volunteer doctors who have no known FtP concerns. Doctors volunteer to take a knowledge test and an Observed Structured Clinical Examination (OSCE) in their relevant specialty. The written paper consists of a single best answer (SBA) knowledge test which is machine marked. The OSCE is marked by trained assessors using a generic domain-based mark scheme of 'acceptable', 'cause for concern' and 'unacceptable'. In the self-assessment literature cited previously,^{6–13} the tests that participants were given differed in three key ways from the tests used in the present study. They had been designed specifically for research purposes to discriminate excellent from poor performance, and the results had no real-world consequences. This was not the case in the present study. Participants were assessed on tests that were established for the purpose of FtP investigations, they are tailored to assess the expected minimum level of competence of a practising doctor and there were potentially real-world consequences if an

individual's performance failed to meet the minimum standard.

The purpose of this study was to investigate how well doctors think they perform on the GMC's ToC pilot examinations. In particular, we wondered whether the established differences in the literature between high and low performers would emerge; do those who perform well underestimate their performance and do those who perform less well overestimate? We were also interested in investigating whether differences in self-estimates existed by gender, ethnic background and Primary Medical Qualification (PMQ) region.

METHODS

The study participants explicitly consented to their data being used anonymously for research purposes.

Design

This was a cross-sectional study using a questionnaire method.

Sample

Doctors who volunteered to take a GMC ToC pilot examination between June 2011 and July 2012 were invited to participate in this study. Volunteers for the pilot examinations were recruited through advertisement in medical journals, specialty-specific newsletters and word of mouth. The study sample included doctors who worked in paediatrics, child psychiatry, anaesthetics, old age psychiatry, forensic psychiatry, general medicine, emergency medicine, orthopaedics, general practice, obstetrics and gynaecology, surgery, cardiology, radiology and care of the elderly. They ranged from foundation year two to consultant level.

Materials

A study-specific questionnaire was designed to obtain participants' estimated knowledge test and OSCE scores (see online supplementary appendix 1). The questionnaire asked where the participants thought they ranked in comparison with other doctors who completed the examinations on the same day and in comparison with all doctors who were eligible to sit the GMC ToC pilot examinations as well as an estimation of their total knowledge test and OSCE scores.

Outcome measures

We compared participants' self-estimated and actual total scores on the knowledge test and OSCE. The knowledge tests consisted of 120 specialty-specific items in a SBA format with a maximum score of 120. The OSCE included 12 specialty-specific stations. Each station was scored by a trained assessor, who was usually a consultant in the relevant specialty, or a clinical skills nurse. The maximum score for the OSCE was 480.

Procedure

Doctors who volunteered to take a GMC pilot examination between June 2011 and July 2012 were invited to participate in this study. Once volunteer doctors had completed the knowledge test and OSCE, the study questionnaire was distributed by GM who was a facilitator at the piloting events. Doctors were briefed about the purpose of the study and how their data would be used. They were assured that the completion of the questionnaire was voluntary and would only take 5–10 min.

Analyses

Actual and estimated examination scores were compared using SPSS for windows V.19. We split estimated and actual scores into tertiles (top, middle and bottom) to see whether differences existed between high and low performers. Results were analysed using descriptive statistics, correlations, t tests and analysis of variances (ANOVAs).

RESULTS

Between June 2011 and July 2012, 689 doctors volunteered to take a pilot ToC and 524 of them participated in the present study (76% rate of participation). During this period, most of them were junior and middle grade doctors who qualified in the UK. As compared with all doctors on the 2011 List of Registered Medical Practitioners (LRMP),¹⁸ men were under-represented and women were over-represented (table 1). There were also a higher proportion of Asian/Asian British doctors in this study, and overseas trained doctors were under-represented (table 1).

General patterns

Overall, participants were more accurate in predicting their total knowledge test score than their OSCE score. There was a moderately strong positive relationship between the difference in estimated and actual

knowledge test scores with actual scores; $r=0.43$, $p<0.01$. There was a roughly equal distribution of participants who overestimated and underestimated their knowledge test scores (figure 1). Those who overestimated (negative numbers on the y axis) tended to score lower than those who underestimated their knowledge test scores. There was no significant difference between the estimated and actual scores on the knowledge test; $t(521)=-1.33$, $p=0.19$.

There was also an association between difference in actual and estimated scores for the OSCE; $r=0.33$, $p<0.01$. The vast majority of doctors underestimated their OSCE performance (figure 2). The few who overestimated their OSCE scores performed less well than those who underestimated their OSCE performance (figure 2). There was a significant difference between the estimated and actual total OSCE scores; $t(520)=-37.76$, $p<0.01$.

Differences between high and lower performers on their estimated and actual examination scores

There were significant differences between high and lower performers on their estimated knowledge test performance. The highest performers significantly underestimated their knowledge test scores by an average of 8 marks ($t(196)=-7.70$, $p<0.01$) and the lower performers significantly overestimated by an average of 7 marks ($t(172)=6.09$, $p<0.01$).

Both high and lower performers significantly underestimated their OSCE performance; $t(180)=-26.28$, $p<0.01$ and $t(172)=-16.20$, $p<0.01$, respectively. Those in the top percentile underestimated their OSCE performance to the greatest extent.

Gender differences on estimated and actual examination scores

Men predicted a higher knowledge test score than did women, with mean estimates of 79 (SD 15) and 74 (SD 14), respectively. Levene's test confirmed there was homogeneity of variance and the t test revealed that this gender

Table 1 Demographic characteristics of sample compared with demographics of doctors on 2011 LRMP

Variable	Levels	Total in this study (N=524)	Total on LRMP 2011 (N=245 903)
Gender	Male	231 (44.1%)	141 369 (57%)
	Female	293 (55.9%)	104 534 (43%)
Ethnicity	White	266 (50.8%)	118 822 (48%)
	Black/black British	17 (3.2%)	6812 (2.8%)
	Asian/Asian British	147 (28.1%)	46 664 (4.3%)
	Mixed	25 (4.8%)	3643 (1.5%)
	Other ethnic groups	20 (3.8%)	9002 (3.7%)
	Not stated (includes prefer not to say)	49 (9.4%)	60 960 (25%)
	Primary Medical Qualification region	UK	408 (77.9%)
	European Union country	22 (4.2%)	24 031 (10%)
	Non-EU country	94 (17.9%)	66 608 (27%)

LRMP, List of Registered Medical Practitioners.

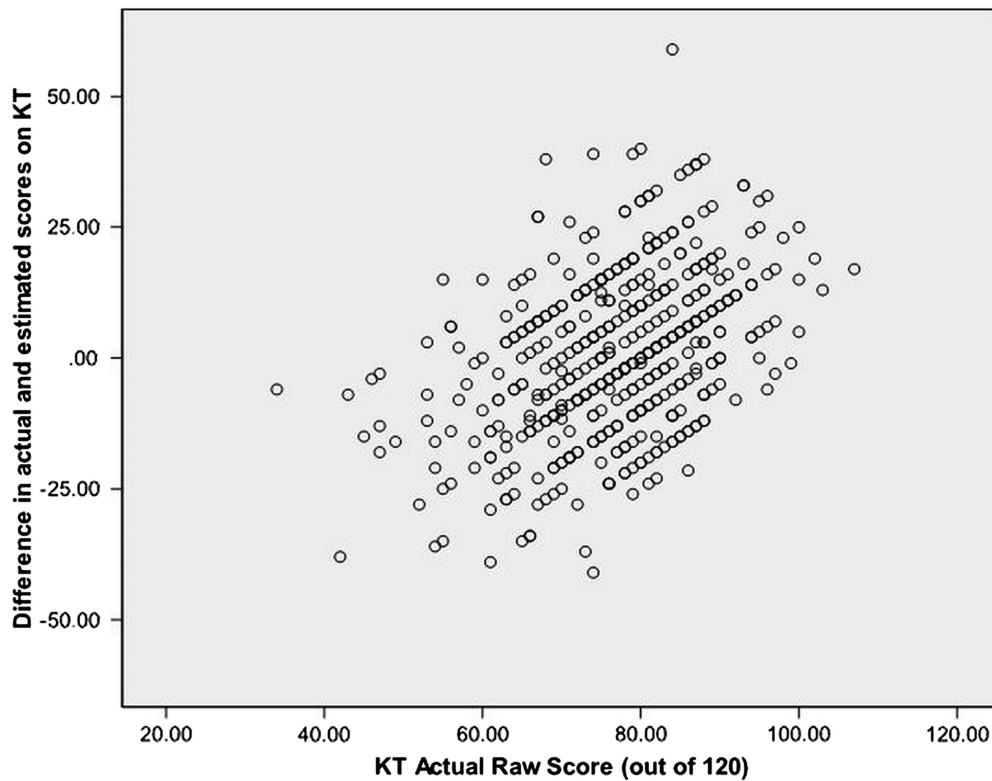


Figure 1 Scatterplot of difference in actual and estimated knowledge test (KT) scores against actual knowledge test scores.

difference in mean estimates was significant; $t(522)=4.27$, $p<0.01$. Women performed slightly better on the knowledge test than did men, but their mean scores were not

significantly different; 77 (SD 11) and 76 (SD 10), respectively. Men also predicted a higher overall OSCE performance than did women with mean estimates of 329 (SD 59)

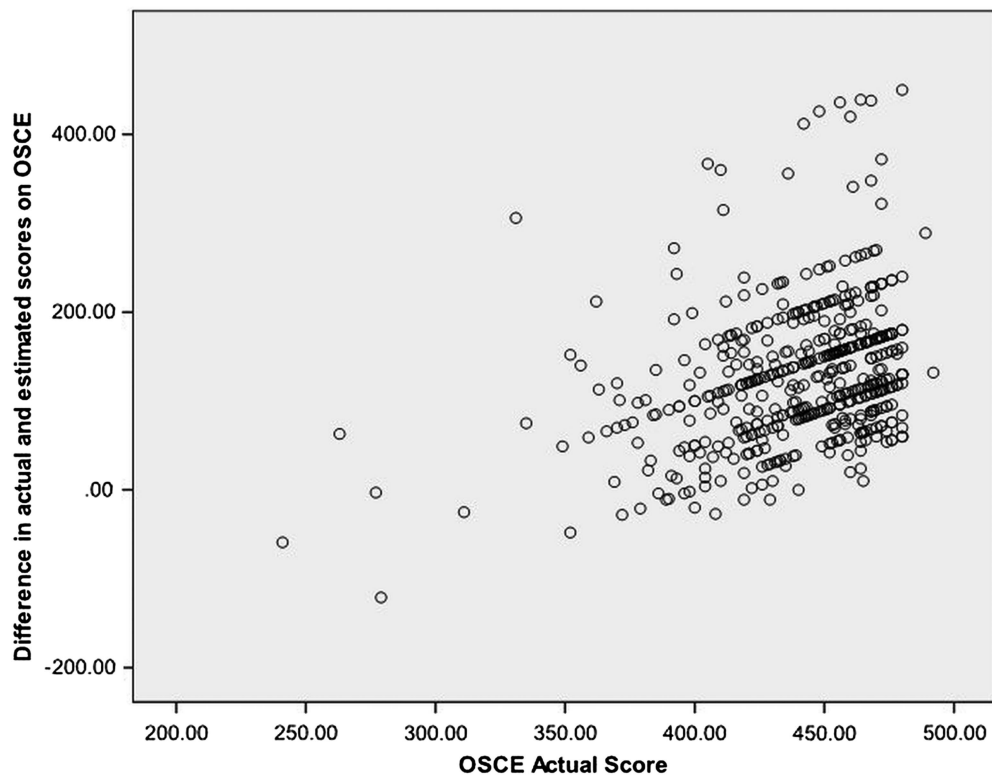


Figure 2 Scatterplot of difference in actual and estimated Observed Structured Clinical Examination (OSCE) scores against actual OSCE scores.

and 300 (SD 81), respectively. This was a significant difference; $t(516)=4.65$, $p<0.01$. Women outperformed men on OSCE performance and the difference was significant; $t(482)=-2.82$, $p<0.01$. A three-way ANOVA showed that there was a significant effect of gender on estimated knowledge test ($F(1508)=4.62$, $p=0.03$) and OSCE performance ($F(1505)=11.74$, $p<0.01$). This means that men, irrespective of their ethnicity or PMQ region, were more likely than women to overestimate their performance on both examinations.

Ethnic differences on estimated and actual examination scores

There were no significant differences in estimated examination scores between doctors of different ethnic backgrounds. However, there was a tendency for Asian/Asian British doctors to overestimate their knowledge test performance to a greater extent than their non-Asian peers ($M=78$, $SD=15$). The highest estimated OSCE performance came from white doctors ($M=317$, $SD=73$) and 'other' ethnic groups ($M=316$, $SD=59$).

A one-way ANOVA confirmed that there were significant differences by ethnic background on actual knowledge test performance ($F(5516)=4.46$, $p<0.01$) and OSCE performance ($F(5518)=11.48$, $p<0.01$). Fisher's least significant difference test was used to explore where these differences occurred. Doctors who did not specify their ethnicity performed highest on the knowledge test (mean=78, $SD=11$). Asian/Asian British doctors scored lowest on the knowledge test, particularly in comparison with white doctors ($p<0.01$) and those who did not state their ethnicity ($p=0.02$). Table 2 shows that white doctors outperformed all other ethnic groups on OSCE performance.

Differences by PMQ region on estimated and actual examination scores

The majority of doctors gained their PMQ from the UK (78%). Of the non-UK trained doctors, 18% were from a non-European Union (EU) country and 4% were from an EU country. There were significant differences in estimated knowledge test performance between doctors of

different PMQ regions. Non-UK trained doctors estimated significantly higher than UK-trained doctors ($F(2521)=6.06$, $p<0.01$). However, there were no actual differences in knowledge test performance between doctors of different PMQ regions ($F(2519)=2.28$, $p=0.10$). A reverse pattern was true of OSCE performance. Estimated OSCE performance did not differ by PMQ region, although EU-trained doctors tended to make the highest estimates. Actual OSCE performance did significantly differ, with UK-trained doctors outperforming non-UK trained doctors ($F(5521)=37.96$, $p<0.01$).

DISCUSSION

Principal findings

In general, participants performed well on the knowledge test and OSCE (usually 70% and above), so the number of actual low performers was small. They were more accurate in predicting their knowledge test performance than OSCE performance. Differences in predictions between high and lower performers were found on the knowledge test but not on the OSCE. In keeping with previous literature,⁶⁻¹³ high performers significantly underestimated their knowledge test performance while lower performers significantly overestimated. Most doctors significantly underestimated their OSCE performance irrespective of how well they actually did. Differences between estimated and actual performance were apparent between men and women. On both examinations, women's estimated performance was lower than men's although women's actual performance was better than men, particularly on OSCE performance. Estimated performance on both examinations did not significantly differ between ethnicities, but there was a tendency for Asian/Asian British participants to estimate slightly higher than other groups. Actual performance for both examinations did significantly differ by ethnic group. Doctors of white and unspecified ethnicity performed highest on the knowledge test, while white doctors outperformed all others on OSCE performance. UK-trained doctors outperformed overseas-trained doctors on OSCE performance but there were no differences by PMQ region on knowledge test performance.

Findings in relation to literature

Our results are in line with the literature that demonstrates the limited ability people have, including doctors, to accurately self-assess their performance.^{5-13 19} Furthermore, this study provides support for previously reported patterns between high and low performers.^{7 13 20-22} There was a tendency for doctors who performed particularly well on the knowledge test to significantly underestimate their score while lower performers significantly overestimated their score. However, OSCE results did not support this pattern. Perhaps, it was easier for doctors to predict their own performance on a machine-marked test of knowledge rather than on a practical skills test that is marked by an

Table 2 Differences in OSCE performance between white and other ethnic groups

Ethnicity	OSCE performance	Significance
White ($M=451$, $SD=29$)		
Not stated	$M=440$, $SD=23$	$p=0.04$
Other	$M=439$, $SD=27$	$p=0.12^*$
Mixed	$M=433$, $SD=39$	$p=0.08^*$
Black/black British	$M=428$, $SD=33$	$p=0.03$
Asian/Asian British	$M=427$, $SD=37$	$p<0.01$

*Non-significant difference.
OSCE, Observed Structured Clinical Examination.

assessor. Furthermore, it is likely that they were unfamiliar with tests that are designed to assess minimum competence and may have assumed the threshold for good performance to be higher than it actually was. From a previous study that we recently conducted, we know that, many of them in this cohort of doctors volunteered to sit a ToC in preparation for their forthcoming postgraduate examinations. Perhaps, a lack of confidence explains the underestimation of OSCE scores that most doctors showed. Alternatively, some authors share the view that people will always be unable to accurately estimate their performance and that this is a methodologically flawed approach to measuring peoples' self-perception.¹⁴⁻¹⁷ While actual performance is poorly correlated with self-ratings (score prediction), there is evidence to suggest it better correlates with behavioural measures. Several studies have shown that when behavioural measures are used, people demonstrate better awareness of their own performance. Psychology students showed awareness of the limits of their knowledge by spending longer time on questions they were unsure about and avoiding answering questions they knew they would get incorrect.¹⁵ A study with medical students who took a qualifying examination reported similar findings.¹⁷ Candidates' self-monitoring was measured according to time taken to respond to each question, the number of questions flagged for further consideration and the likelihood of changing their initial answer. This study found that high performers demonstrated better self-monitoring than poorer performers on the examination.¹⁷ Following this evidence, there are recommendations to pursue this line of research approach instead of asking people to estimate their own examination scores.¹⁴⁻¹⁷

The gender and ethnic differences found in this study support that of previous findings to an extent. Women tend to underestimate and men tend to overestimate their performance in medicine.^{21 23-25} Our findings supported this pattern on knowledge test performance but not on OSCE performance. White medical students consistently outperform non-white medical students in the UK,²⁶ the USA²⁷⁻³¹ and in other English speaking countries.^{32 33} In this study, white doctors performed substantially higher than non-white doctors on the OSCE but not on the knowledge test. Furthermore, there was an interaction between ethnicity and PMQ region. OSCE performance was higher in white UK-trained doctors than in white non-UK-trained as well as non-white UK-trained doctors.

Implications of findings

Overall, most doctors did not appear to have an inflated view of their examination performance. It is reasonable to assume that doctors who are not overly confident are likely to exercise more caution in their clinical practice than those who are overconfident. However, roughly half of the sample did overestimate their knowledge test performance. This is potentially a problem as overconfidence in doctors is associated with poor clinical judgement and decision-

making.^{34 35} Furthermore, research has shown that overconfidence in medicine is more likely to be a male rather than female characteristic.^{21 23} Women are more likely to perceive themselves and be perceived by others as less confident in clinical knowledge and skills.^{21 23} This pattern was found in our study, despite women outperforming men on the knowledge test and OSCE. In practice, lack of confidence may disadvantage female doctors with patient interaction and career progression.²³ A common pattern of ethnic differences was also found with white doctors outperforming non-white doctors on the OSCE. The reasons for this performance gap are unclear but cross-cultural differences in communication styles may explain some of the variations in performance on OSCE type examinations; assessors may also have been influenced by ethnic stereotypes.³⁶ This performance gap is in line with the recent controversy around higher rates of failure among international medical graduates taking the clinical skills assessment portion of the Medical Royal College of General Practitioners (MRCGP) examination.³⁷ Further research is necessary to understand why these ethnic differences persist in medicine and what can be done to reduce this discrepancy.^{26 36 38}

Medical education could facilitate the development of doctors' accurate self-perception by including formal training on the biases that affect the self-perception of all individuals.^{16 34 35} Doctors would learn about the inherent heuristics they are likely to use when reflecting on strengths and weaknesses of their performance.^{34 35} Medical educators should also establish how feedback can be delivered in a way that is likely to be internalised to encourage the necessary behavioural changes. One potential benefit of revalidation is that it will enforce doctors to reflect on their clinical knowledge and practice as well as identify areas for improvements.¹ This process may prove to be a good opportunity for doctors to become more self-aware. This remains to be seen in future research once the current round of revalidation ends in 2016. In practice, the revalidation process is interested in doctors' awareness and monitoring of the limits of their knowledge and clinical skills, rather than their ability to accurately predict their assessment scores. Therefore, an understanding of the universal biases that affect self-perception, coupled with appropriate behaviour changes in response to feedback, is likely to improve doctors' self-perception and capacity to self-monitor.

Strengths and weaknesses of the study

This study lends further support to the literature that suggests doctors have limited ability to estimate their examination performance, even when the examinations are in a familiar format (SBA and OSCE). It is one of the first studies to look at how well a group of doctors think they perform on a set of voluntary examinations that have potentially significant consequences. The large sample means that it has greater power than previous similar studies. The tests included were part of a

validation process for the GMC's FtP procedures. Therefore, the study has practical relevance, as demonstrated by doctors seeking feedback on their performance and commenting that they used the process as a preparation for future examinations. There is concern that perhaps only those who thought they had performed well on the examinations would have participated in this study, thus introducing selection bias into the sample. However, most of the doctors who were invited to participate did so (76%), and we know from the results that most doctors did not have an inflated view of their examination performances. A limitation of this study is that no measures of behaviour were included that could have demonstrated the extent to which doctors monitor their performance in a given clinical situation. We recognise the value in this alternative approach for extending the present findings in future research. However, in the case of this study, the doctors primarily volunteered to take a ToC than to be in a research study. For this reason, a questionnaire asking for self-estimated scores on the examinations they had just taken was a feasible way for us to obtain data on this topic. The results are not necessarily generalisable to all doctors and may have differed had there been equal numbers of doctors by different ethnic background, PMQ regions and seniority. Finally, the majority of doctors performed well on both examinations, including those whose scores were in the bottom percentiles. Therefore, patterns between high and lower performers should be interpreted with caution. Those who performed less well than the majority may have struggled because of the type of examination material. However, it is likely that the few who performed lower in comparison with their peers who took the same examination material were actual low performers.

CONCLUSIONS AND FUTURE DIRECTIONS

Doctors were more accurate in predicting their knowledge test performance than OSCE performance. High and lower performers self-estimated differently on an objective test of knowledge but almost everyone underestimated their performance on a practical skills test (OSCE). Estimated and actual performance differed by gender and ethnicity but less so by where a doctor had gained their PMQ. A follow-up to this study may wish to explore in more depth how doctors come to assign themselves a particular score and their reasoning underpinning this judgement. Such data may lead to further understanding of why high performers tend to underestimate their own performance and low performers overestimate. Anecdotally we know that doctors undergoing FtP investigation often lack sufficient recognition of their problems. Further study on how poor performers in particular can successfully alter their self-perception as a first step towards remediation is warranted. It will be interesting to monitor the impact revalidation has on

the number of complaints to the GMC and whether this formal exercise in self-reflection affords an opportunity for borderline problematic doctors to rectify their deficiencies.

Acknowledgements The authors are grateful to Dr Henry Potts for his statistical support and help with interpreting results.

Contributors LM analysed the results and wrote the manuscript. AS contributed intellectually to the study and helped revise the manuscript drafts. GM designed the research questionnaire and helped recruit participants. YK recruited participants and entered the data. JD oversaw the study and provided intellectual contribution at all phases. All authors provided feedback on manuscript drafts and approved its final version.

Funding This work was supported by the General Medical Council.

Competing interests None.

Ethics approval Received written confirmation from University College London's Research Ethics Committee in October 2008 that the study was exempt from ethical approval.

Provenance and peer review Not commissioned, externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

1. General Medical Council. Ready for revalidation: supporting information for appraisal and revalidation. 2012.
2. General Medical Council. Good Medical Practice. 2012.
3. Hays R, Jolly B, Caldon LJM, *et al*. Is insight important? Measuring capacity to change performance. *Med Educ* 2002;36:965–71.
4. Cross P. Not can but will college teaching be improved? *New Dir Higher Educ* 1977;17:1–15.
5. Dunning D, Heath C, Suls J. Flawed self-assessment: implications for health, education, and the workplace. *Psychol Sci Public Interest* 2004;5:69–106.
6. Ehrlinger J, Johnson K, Banner M, *et al*. Why the unskilled are unaware: further explorations of (absent) self-insight among the incompetent. *Organ Behav Hum Decis Process* 2008;105:98–121.
7. 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;7:1121–34.
8. Marottoli R, Richardson E. Confidence in, and self-rating of, driving ability among older drivers. *Accid Anal Prev* 1998;30:331–6.
9. Rutter D, Quine L, Albery I. Perceptions of risk in motorcyclists: unrealistic optimism, relative realism and predictions of behaviour. *Br J Psychol* 1998;89:681–96.
10. Zenger T. Why do employers only reward extreme performance? Examining the relationships among performance, pay, and turnover. *Adm Sci Q* 1992;37:198–219.
11. Tracey J, Arroll B, Barham P, *et al*. The validity of general practitioners' self assessment of knowledge: cross sectional study. *BMJ* 1997;315:1426.
12. 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:S87–9.
13. Davis D, Mazmanian PE, Fordis M, *et al*. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *J Am Med Assoc* 2006;296:1094–102.
14. Ward M, Gruppen L, Regehr G. Measuring self-assessment: current state of the art. *Adv Health Sci Educ* 2002;7:63–80.
15. Eva K, Regehr G. Knowing when to look it up: a new conception of self-assessment ability. *Acad Med* 2007;82(Suppl 10):S81–4.
16. Eva K, Regehr G. I'll never play professional football and other fallacies of self-assessment. *J Contin Educ Health Prof* 2008;28:14–19.
17. McConnell M, Regehr G, Wood T, *et al*. Self-monitoring and its relationship to medical knowledge. *Adv Health Sci Educ* 2012;17:311–23.

18. General Medical Council. The State of Medical Education and Practice in the UK. 2012.
19. Dunning D, Johnson K, Ehrlinger J, *et al*. Why people fail to recognize their own incompetence. *Curr Dir Psychol Sci* 2003;12:83–7.
20. Barnsley L, Lyon P, Ralston S, *et al*. Clinical skills in junior medical officers: a comparison of self-reported confidence and observed competence. *Med Educ* 2004;38:358–67.
21. Blanch-Hartigan D. Medical students' self-assessment of performance: results from three meta-analyses. *Patient Educ Couns* 2011;84:3–9.
22. Tousignant M, Desmarchais J. Accuracy of student self-assessment ability compared to their own performance in a problem-based learning medical program: a correlation study. *Adv Health Sci* 2002;7:19–27.
23. Blanch D, Hall J, Roter D, *et al*. Medical student gender and issues of confidence. *Patient Educ Couns* 2008;72:374–81.
24. Minter R, Gruppen L, Napolitano KS, *et al*. Gender differences in the self-assessment of surgical residents. *Am J Surg* 2005;189:647–50.
25. Rees C, Shepherd M. Students' and assessors' attitudes towards students' self-assessment of their personal and professional behaviours. *Med Educ* 2005;39:30–9.
26. Woolf K, Potts H, *et al*. Ethnicity and academic performance in UK trained doctors and medical students: systematic review and meta-analysis. *BMJ* 2011;342:901.
27. Colliver J, Swartz M, Robbs R. The effect of examinee and patient ethnicity in clinical-skills assessment with standardized patients. *Adv Health Sci Educ* 2001;6:5–13.
28. Dawson B, Iwamoto C, Ross L, *et al*. Performance on the national board of medical examiners. Part 1 examination by men and women of different race and ethnicity. *J Am Med Assoc* 1994;272:674–9.
29. Koenig J, Sireci S, Wiley A. Evaluating the predictive validity of MCAT scores across diverse applicant groups. *Acad Med* 1998;73:1095–106.
30. Veloski J, Callahan C, Xu G, *et al*. Prediction of students' performances on licensing examinations using age, race, sex, undergraduate GPAs, and MCAT scores. *Acad Med* 2000;75:S28–30.
31. Xu G, Veloski J, Hojat M, *et al*. Longitudinal comparison of the academic performances of Asian-American and white medical students. *Acad Med* 1993;68:82–6.
32. Kay-Lambkin F, Pearson S, Rolfe I. The influence of admissions variables on first year medical school performance: a study from Newcastle University, Australia. *Med Educ* 2002;36:154–9.
33. Liddell M, Koritsas S. Effect of medical students' ethnicity on their attitudes towards consultation skills and final year examination performance. *Med Educ* 2004;38:187–98.
34. Croskerry P. Achieving quality in clinical decision making: cognitive strategies and detection of bias. *Acad Emerg Med* 2002;9:1184–204.
35. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med* 2003;78:775–80.
36. Woolf K, Cave J, Greenhalgh T, *et al*. Ethnic stereotypes and the underachievement of UK medical students from ethnic minorities: qualitative study. *BMJ* 2008;337:a1220.
37. Kaffash J. RCGP faces legal threat over international GP trainee failure rates, in Pulse. 2012.
38. Woolf K, McManus IC, Potts HWW, *et al*. The mediators of minority ethnic underperformance in final medical school examinations. *Br J Educ Psychol* 2011;83:135–59.