

RESEARCH

PLAB and UK graduates' performance on MRCP(UK) and MRCGP examinations: data linkage study

 OPEN ACCESSI C McManus *professor of psychology and medical education*^{1,2}, Richard Wakeford *life fellow*³¹UCL Medical School, University College London, London WC1E 6BT, UK; ²Research Department of Clinical, Educational and Health Psychology, University College London; ³Hughes Hall, University of Cambridge, Cambridge CB1 2EW, UK

Abstract

Objectives To assess whether international medical graduates passing the two examinations set by the Professional and Linguistic Assessments Board (PLAB1 and PLAB2) of the General Medical Council (GMC) are equivalent to UK graduates at the end of the first foundation year of medical training (F1), as the GMC requires, and if not, to assess what changes in the PLAB pass marks might produce equivalence.

Design Data linkage of GMC PLAB performance data with data from the Royal Colleges of Physicians and the Royal College of General Practitioners on performance of PLAB graduates and UK graduates at the MRCP(UK) and MRCGP examinations.

Setting Doctors in training for internal medicine or general practice in the United Kingdom.

Participants 7829, 5135, and 4387 PLAB graduates on their first attempt at MRCP(UK) Part 1, Part 2, and PACES assessments from 2001 to 2012 compared with 18 532, 14 094, and 14 376 UK graduates taking the same assessments; 3160 PLAB1 graduates making their first attempt at the MRCGP AKT during 2007-12 compared with 14 235 UK graduates; and 1411 PLAB2 graduates making their first attempt at the MRCGP CSA during 2010-12 compared with 6935 UK graduates.

Main outcome measures Performance at MRCP(UK) Part 1, Part 2, and PACES assessments, and MRCGP AKT and CSA assessments in relation to performance on PLAB1 and PLAB2 assessments, as well as to International English Language Testing System (IELTS) scores. MRCP(UK), MRCGP, and PLAB results were analysed as marks relative to the pass mark at the first attempt.

Results PLAB1 marks were a valid predictor of MRCP(UK) Part 1, MRCP(UK) Part 2, and MRCGP AKT ($r=0.521, 0.390, \text{ and } 0.490$; all $P<0.001$). PLAB2 marks correlated with MRCP(UK) PACES and MRCGP CSA ($r=0.274, 0.321$; both $P<0.001$). PLAB graduates had significantly lower MRCGP(UK) and MRCGP assessments (Glass's $\Delta=0.94, 0.91, 1.40, 1.01, \text{ and } 1.82$ for MRCP(UK) Part 1, Part 2, and PACES and MRCGP AKT and CSA), and were more likely to fail assessments and to progress more slowly than UK medical graduates. IELTS scores correlated significantly with later performance, multiple regression showing that the effect of PLAB1 ($\beta=0.496$) was much stronger than the effect of IELTS ($\beta=0.086$). Changes to PLAB pass marks that would

result in international medical graduate and UK medical graduate equivalence were assessed in two ways. Method 1 adjusted PLAB pass marks to equate median performance of PLAB and UK graduates. Method 2 divided PLAB graduates into 12 equally spaced groups according to PLAB performance, and compared these with mean performance of graduates from individual UK medical schools, assessing which PLAB groups were equivalent in MRCP(UK) and MRCGP performance to UK graduates. The two methods produced similar results. To produce equivalent performance on the MRCP and MRCGP examinations, the pass mark for PLAB1 would require raising by about 27 marks (13%) and for PLAB2 by about 15-16 marks (20%) above the present standard.

Conclusions PLAB is a valid assessment of medical knowledge and clinical skills, correlating well with performance at MRCP(UK) and MRCGP. PLAB graduates' knowledge and skills at MRCP(UK) and MRCGP are over one standard deviation below those of UK graduates, although differences in training quality cannot be taken into account. Equivalent performance in MRCGP(UK) and MRCGP would occur if the pass marks of PLAB1 and PLAB2 were raised considerably, but that would also reduce the pass rate, with implications for medical workforce planning. Increasing IELTS requirements would have less impact on equivalence than raising PLAB pass marks.

Introduction

International medical graduates who wish to practise medicine in the UK can be accepted onto the List of Registered Medical Practitioners of the General Medical Council (GMC) by passing the two examinations set by the GMC's Professional and Linguistic Assessments Board (PLAB). International medical graduates usually possess qualifications from outside the European Economic Area (EEA), as doctors with an EEA medical qualification who have EU rights can normally be registered under reciprocal arrangements. Figure 1↓ provides a synopsis of the training and assessment undertaken in the UK by international medical graduates and by UK medical graduates.

PLAB Part 1 is a multiple choice assessment of medical knowledge in four domains (context, diagnosis, investigation, management), and PLAB Part 2, which is an objective structured clinical examination, also assesses in four domains (communication, history taking, examination, practical skills). A current pre-condition for taking PLAB is that international medical graduates have within the previous two years achieved an acceptable level at IELTS (International English Language Testing System¹) with a score of at least 7 in each of its four domains (listening, reading, writing, speaking). In the five years from 2008-12 an average of 1281 international medical graduates per year passed PLAB, and in the same period an annual average of 6720 UK graduates fully registered with the General Medical Council (GMC, personal communication). PLAB graduates are similar in number to the output of four or five medium sized UK medical schools.

The desired standard for the PLAB exams has been consistently stated since the introduction of the assessment in 1975, when it was known as the TRAB (Temporary Registration Assessment Board) assessment.²⁻⁴ In the 2003 review of PLAB the standard was justified and summarised thus: "Council has agreed that ... [it] would be inequitable to expect UK-trained doctors and international medical graduates to satisfy different standards to obtain full registration. For these reasons we have concluded that the standard of the test should be that of doctors completing the end of Foundation Year 1"(Para 15).⁴

In 2011 the GMC set up a working party to review the PLAB examinations once again. Included within the remit was an assessment of whether "the knowledge and skills demonstrated by a pass in the PLAB test continue to be equivalent to those demonstrated by successful completion of [Foundation Year 1] training."

In addition the working party was asked "to examine whether international medical graduates granted full registration following a successful pass in the PLAB test are more or less likely than other cohorts of doctors to experience difficulties in medical practice in the UK" by "examining any evidence of disparity between the success rates of UK medical graduates and international medical graduates in postgraduate examinations and assessments."⁵

As a part of addressing these questions the Working Party commissioned two sets of primary research, and the present study is one.

International medical graduates and the MRCGP

Although the PLAB Working Party had been considering the performance of international medical graduates before then, the performance of international medical graduates in postgraduate medical examinations came under intense scrutiny in 2013 in relation to pass rates of international medical graduates in the MRCGP. The examination is in two parts, the AKT (Applied Knowledge Test, a multiple choice examination) and the CSA (Clinical Skills Assessment, a 13-station simulated surgery in objective structured clinical examination format). In February 2013, leave for a judicial review into differential pass rates of international medical graduates at the CSA was sought by the British Association of Physicians of Indian Origin and agreed to by the administrative court in July 2013 (and took place in April 2014). In April 2013 the GMC also set up an independent enquiry into the MRCGP, the ensuing "Esmail and Roberts report" being published in September 2013,⁶ along with a parallel article in the *BMJ*.⁷ That report examined the performance of 5095 doctors who had taken MRCGP exams

and for whom ethnicity was known (2663 being white and 2432 being "black and minority ethnic"). Of these doctors, 1310 candidates were international medical graduates, 3644 were UK graduates, and 141 were EEA graduates, with most international medical graduates being classified as black and minority ethnic.

Esmail and Roberts reported that international medical graduates were 14.7 times more likely to fail the CSA than UK graduates after "correcting for age, gender and performance at AKT," and 2.9 times more likely to fail the AKT.⁷ In addition, among UK graduates, black and minority ethnic candidates were 3.5 times more likely to fail the CSA than white graduates. The report to the GMC concluded that differences on the machine-marked AKT were "difficult to attribute to ... bias" and that "the reasons for the differential pass rates are likely to be complex" (p 14),⁶ and were consistent with differences reported more generally in medical examinations⁸ at both the postgraduate and undergraduate level (for which many possible explanations have been tested^{9 10}).

Esmail and Roberts posited a number of possible reasons for the lower performance of international medical graduates in the CSA, including differences in preparedness for an assessment, "which is not a culturally neutral examination and nor is it intended to be" (p 15).⁶ The format of the CSA examination itself was not felt to be a problem, being "based on a well-established pedagogy." However, it was noted that "the nature of the examination is such that it is open to subjective bias" on the part perhaps of examiners or of simulated patients, although no statistical evidence was presented. A recommendation of the Esmail and Roberts report was that "further research should be commissioned ... to investigate how black and minority ethnic standardised patients and black and minority ethnic examiners score candidate physicians who are racially and ethnically concordant and compare that to how non-concordant standardised patients and examiners score the black and minority ethnic candidates" (p 19). Group analyses of examiner and candidate concordance for ethnicity in MRCGP by one of us find little evidence of bias,¹¹ and are consistent with similar analyses of MRCP(UK) at the group level¹² and the individual examiner level.¹³ Despite Esmail and Roberts' claim that "subjective bias owing to racial discrimination cannot be excluded,"⁷ it seems unlikely from our empirical analyses¹¹⁻¹³ that racial discrimination is an explanation for differential performance by international medical graduates in exams such as MRCGP and MRCP(UK).

The lower performance of international medical graduates in the MRCGP examination is not unique to that exam, although data from other postgraduate examinations are less easy to interpret as some or many international medical graduates are not UK trainees or are not even registered in the UK. Bearing that in mind, the MRCP(UK) has published data for some years from which it is clear that international medical graduates perform less well.¹⁴ Lower international medical graduate performance has also been reported in the MRCPsych examination,¹⁵ and it is also clear that international medical graduates perform less well in the MRCOG examination^{16 17} and in the assessments towards MRCPCH.¹⁸

Equivalence

A central, but difficult, concept which is present within the remit of the GMC working party is the concept of equivalence—the term being used explicitly in one remit ("continue to be equivalent"), which we will call "entry equivalence," and implicitly in the other ("more or less likely ... to experience difficulties"), which we will call "outcome equivalence,"

equivalence being neither more nor less likely to experience difficulties. Problems in defining equivalence also occur with the Certificate of Eligibility for Specialist Registration and Certificate of Eligibility for GP Registration, which are alternative routes for international medical graduates (and some UK graduates) to enter the specialist or GP registers.¹⁹

Within medicine the concept of equivalence testing has been used since the 1980s in clinical pharmacology to assess whether two compounds are sufficiently similar to be considered equivalent²⁰ (and the methodology is also used elsewhere²¹). Equivalence testing typically considers a single parameter, such as the mean or the peak level of a drug. Although a mean can describe a distribution, it is not the only important parameter. The abilities of UK and PLAB graduates form distributions, with some graduates being excellent and others being barely acceptable. The phrase “equivalent to ... doctors who have successfully completed F1” for defining entry equivalence is unclear. Although it could mean that means or medians should be equivalent, it might also be interpreted, since PLAB is a qualifying examination, that all PLAB graduates should be at least as good as, say, the worst UK graduate on the register.

The main concern of the present study is in assessing outcome equivalence in relation to MRCP(UK) and MRCGP, and we will compare the median performance of PLAB and UK graduates, and we will also compare PLAB graduates who pass the exam at different levels with the average performance of graduates from different UK medical schools.

Direct and indirect comparison

Evaluating the equivalence of different assessments is never straightforward unless either there are two groups of individuals taking the same assessment²² or there is cross moderation of judgemental methods²³ allowing a direct comparison. UK graduates who have “successfully completed the first year of Foundation Programme training” do not take PLAB (or indeed any other summative assessment at the end of the first foundation year), and PLAB graduates will not have taken UK medical school finals. Neither are there shared questions in PLAB and UK medical school finals (indeed, because UK medical schools run their own final examinations, different items are used in different schools, and standards may differ between UK medical schools²⁴). Direct assessment of the entry equivalence of PLAB is not therefore possible at present. An indirect assessment of entry equivalence could compare groups such as PLAB and UK graduates on some other assessment taken by both groups—an external yardstick. For the present study the yardstick is performance in the MRCGP and MRCP(UK) examinations, and the yardstick of the Annual Review of Competence Progression is analysed in a separate report by a different team.²⁵

The logic of the current study is straightforward: MRCP(UK) and MRCGP exams are taken by both UK and PLAB graduates, and if the UK and PLAB graduates are equivalent in their outcomes then they should perform equally well when they take the MRCP(UK) and MRCGP. The situation is made somewhat more complex as UK and PLAB graduates choose which medical specialty to enter, and they are also selected onto training programmes such as for general practice or for core medical training. Those taking the examinations may not therefore be representative samples, although they are at least complete samples of UK and PLAB graduates taking MRCP(UK) and MRCGP in the years concerned.

Validity

Our analyses can be considered as an exercise in assessing the validity of the PLAB assessments. High stakes examinations have a pass mark (“cut score” or “passing score”), and, although little discussed in the literature, a key question concerns the validity of that pass mark. Kane²⁶ distinguishes clearly between a pass mark and a performance standard, the latter being a measure of adequate performance in the domain to which passing the assessment allows access. For Kane, “Validation ... consists of a demonstration that the proposed passing score can be interpreted as representing an appropriate performance standard.”

Kane distinguishes several types of validity. “Procedural validity” is the appropriateness of the procedures used in standard setting, with poor procedure casting doubt on the validity of a pass mark but good procedure alone being unable to validate a pass mark. “Internal validity” of standard setting assesses examiner agreement on the pass mark, and it alone also cannot validate a pass mark—as Verheggen et al wrote, judgments can be “more reliable, [but] they may [also] be less valid. In other words the judgements would be consistently off the mark” (p 210).²⁷ Kane’s third approach to validity uses external criteria, particularly the “direct, criterion-related approach,” asking how those passing the exam perform at later tasks, whether those who pass well perform better than those who only just pass and whether those only just passing subsequently perform at an acceptable level. That is the approach adopted in this paper, although we refer to it as indirect.

The present study

The study reported here uses record linkage, based on GMC registration number, to assess performance in a large group of PLAB graduates who have gone on to take the MRCP(UK) and/or MRCGP examinations. The analyses presented here differ from those reported by Esmail and Roberts in some important ways. Firstly, the analyses have a large dataset from MRCP(UK), and, secondly, extensive analysis is carried out of PLAB Part 1 examination results (which Esmail and Roberts chose not to study (p 11)⁶). The primary interest is in the extent to which PLAB and UK graduates are equivalent in their subsequent postgraduate performance, with a secondary interest in whether a change in the standard set for the PLAB assessments could result in outcome equivalence. As will be discussed later, it is accepted that other factors might also be of importance in determining differences in international medical graduate and UK medical graduate performance.

Method

The study used data linkage for the main analyses, with data protection and other issues constraining how the linkage took place. Data linkage in the first instance took place at the GMC, to which was sent by the MRCP(UK) and MRCGP office the GMC number, name, date of birth, and place of primary medical qualification of all candidates known to have taken MRCP(UK) or MRCGP. The GMC did not receive either MRCP(UK) or MRCGP examination results themselves. Having identified doctors in those sets who had also taken PLAB, the GMC then sent data files with information on PLAB and IELTS performance to ICM and RW, who separately linked the PLAB and IELTS data with MRCP(UK) performance and MRCGP performance. The fully linked datasets were available to ICM and RW only, on a research basis, each processing data only from their own college, and were not available either to the

GMC or to the Royal Colleges of Physicians or the Royal College of General Practitioners.

Descriptions of the various sets of data

MRCP(UK)—Run by MRCP(UK) Central Office for the Federation of Royal Colleges of Physicians of London, Edinburgh and Glasgow, the exam is in three parts. MRCP Part 1 is a 200-item, “best of five” multiple choice assessment with brief clinical vignettes, which is computer marked. MRCP Part 2 is a 270-item, computer marked, “best of five” assessment with more complex and extensive clinical scenarios. PACES (Practical Assessment of Clinical and Examination Skills) is a modified objective structured clinical examination, with eight encounters, six involving real patients and two involving simulated patients, with two examiners present at each station. The original PACES examination²⁸ changed its format in 2009 to new PACES (nPACES).²⁹ The MRCP(UK) Part 1 and Part 2 examinations have had essentially the same structure since 2001-02,³⁰ although the method of standard setting for both was changed from the Angoff method to statistical equating in 2009 and 2010 respectively. Part 1 can be taken 12 months after graduation. Part 2 and PACES can be taken in any order once Part 1 has been passed.

MRCGP—These examinations, which are run by the Royal College of General Practitioners, are in two parts, the AKT (Applied Knowledge Test, a 200-item, computer displayed and computer marked, multiple choice test with a variety of item types) and the CSA (Clinical Skills Assessment, a 13-station objective structured clinical assessment in the form of a simulated surgery with candidates seeing simulated patients while being assessed by an examiner). The AKT is typically taken during the second year of training, and the CSA is taken during the third and normally final year of training. All candidates are on UK training schemes overseen by postgraduate deaneries; entry to the examinations by others (such as foreign based candidates) is not allowed.

PLAB—PLAB Part 1 (knowledge assessment) is currently a multiple choice, best of five examination with 200 items, of which a small number are removed because of problems in keying or scoring, a typical exam having 197 scored items. The pass mark is set by a variant of the Angoff method and is typically about 125, but has varied in the range 116 to 135. Marks on the four subscales are not reported here. PLAB Part 2 (the clinical assessment) is an objective structured clinical assessment, candidates being assessed on 15 stations, one of which is a non-scoring pilot station. There are four types of station, but subscores will not be considered here. There is a single examiner at each station, and the standardised patients do not take part in marking the assessments. The marking scheme is complex, but has been described elsewhere (www.gmc-uk.org/doctors/plab/borderline_group_scoring_faqs.asp), along with the standard setting method, which is a variant of the borderline group method.

IELTS—The required IELTS level for PLAB has varied over the years but is currently set at a score of 7 on the total score and at all four subscores. Candidates taking PLAB in earlier years may have had lower scores either overall or on subscales. Some PLAB candidates are exempted from the required IELTS level, primarily by demonstrating that their training was at a medical school where the great majority of teaching is in English. Analyses of IELTS are restricted here to the overall score attained.

Mark relative to the pass mark

Pass marks vary from diet to diet of the various examinations, and therefore performance at MRCP(UK), MRCGP, and PLAB is described in terms of mark relative to the pass mark, so that a candidate scoring zero just passes the exam, a candidate with a positive mark has passed the examination with marks to spare, and candidates with a negative mark have failed the examination. We have also carried out analyses based on marks attained at passing, but they are more complex, in particular having skewed, censored distributions, and are less statistically sensitive but give broadly similar results.

Repeated attempts at examinations

Candidates who fail assessments can repeat MRCP(UK), MRCGP, and PLAB. Here we use candidates' marks at first attempt for all analyses, as did Esmail and Roberts. Previous analyses of MRCP(UK) have suggested that mark at the first attempt of taking an examination is the best predictor of future performance³¹ and thus the most accurate measure of ability.

Statistical methods

Statistical analyses used IBM Statistical Package for the Social Sciences v21. Effect sizes are calculated as Glass's delta (Δ), which expresses the performance of PLAB graduates relative to the UK graduates who are regarded as the reference group.

Results

Linkage of the MRCP(UK) and PLAB databases

The database for the current analysis consisted initially of all 65 115 candidates who had taken at least one part of the MRCP(UK) examination between 2001 and 2012. Of these, 37 329 had a GMC number and therefore had at some point worked in the UK. Linkage with the PLAB database identified 9818 PLAB candidates who were also MRCP(UK) candidates. Of the remaining MRCP(UK) candidates, 24 641 had graduated at UK medical schools and are the group to be compared with the PLAB candidates and with whom they should be equivalent. Results at first attempt were not available for all candidates as exams may have been taken outside of the available time window. Marks at first attempt were available for 18 532 UK graduates at Part 1, 14 094 UK graduates at Part 2, and 14 376 UK graduates at PACES, and for 7829 PLAB graduates at Part 1, 5135 PLAB graduates at Part 2, and 4387 PLAB graduates at PACES.

Linkage of the MRCGP and PLAB databases

Two databases were created for the MRCGP and PLAB linkage, one for the AKT and the other for the CSA. Linkage with the PLAB database was carried out by the GMC looking for all PLAB candidates who had a GMC number in the lists of those taking either or both parts of the MRCGP. There were data available on the AKT between 2008 and 2013 for a total of 22 081 candidate attempts, of which 17 395 were first attempts. Of these first attempts, 3160 were for PLAB Part 1 and 3067 for PLAB Part 2, and 2985 had IELTS scores reported. For the current version of the CSA (2010-13), from a total of 11 673 candidate attempts, 8346 were first attempts. Of these, 1411 were for PLAB Part 1, and 1388 for PLAB Part 2, and 1353 had IELTS scores reported.

Representativeness of PLAB graduates taking MRCP(UK) and MRCGP

PLAB graduates taking MRCP(UK) or MRCGP may be different from those who do not take those examinations. Table 1¹ shows performance at PLAB Part 1 and PLAB Part 2 in all doctors who passed PLAB 1 between 4 July 2000 and 13 July 2006 and passed PLAB 2 between 13 June 2001 and 12 January 2007 in relation to whether they had ever taken the MRCGP or MRCP(UK) exams.

PLAB graduates who took MRCP(UK) performed somewhat better on their first attempt at PLAB Part 1, although the effect is small, and they performed a little worse at their first attempt at PLAB Part 2. PLAB graduates who took the MRCGP exams scored somewhat lower on their first attempt at PLAB Part 1 and a little better at their first attempt at PLAB Part 2.

Comparison of UK and PLAB graduates on demographics and progression

Table 2² shows basic descriptive data on demographics and progression for PLAB and UK graduates taking MRCP(UK) and the MRCGP.

For the MRCP(UK), PLAB graduates are more likely to be male and to be from ethnic minorities. UK and PLAB graduates qualify as doctors at similar ages, but PLAB graduates take MRCP(UK) later than UK graduates, not least because they have been taking PLAB Parts 1 and 2 between graduation and taking MRCP(UK) Part 1. PLAB graduates also progress more slowly through MRCP(UK) Parts 1, 2, and PACES (in large part due to having more resits, data not shown).

For the MRCGP, PLAB graduates are more likely than UK graduates to be male and far more likely to be non-white. PLAB graduates are four years older when they take the AKT and six years older when they take the CSA. PLAB graduates are far more likely to resit both the AKT and CSA than UK graduates (mean attempt number in AKT database=1.16 for UK graduates, 1.64 for PLAB graduates, $P<0.001$; mean attempt number in CSA database=1.12 for UK graduates, 2.17 for PLAB graduates, $P<0.001$).

Data on nationality are available only for the candidates taking PLAB, but, as table 2² shows, there is a large group of PLAB candidates who are UK nationals, about 8% (749/9589) for those taking MRCP(UK), and 12% (388/3233) for the MRCGP. The MRCGP candidates who were UK nationals took significantly more attempts to pass PLAB Part 1 than those who were of other nationalities (mean: 1.8 attempts *v* 1.4 attempts, $P<0.001$), first attempt score on PLAB Part 1 was also significantly lower than for non-UK nationals (mean 2.98 *v* 7.21, $P<0.001$), and they also performed less well on the AKT (mean -2.54 *v* 4.67, $P<0.001$), whereas on the CSA they were not statistically different from non-UK nationals (mean -3.45 *v* -4.81, not significant).

Correlation of PLAB results with MRCP(UK) and MRCGP results

If PLAB is a valid assessment of skills relevant to progression during UK postgraduate training then performance on it should relate to performance on subsequent UK postgraduate assessments. Elsewhere, in longitudinal studies of UK graduates, it has been shown that there are strong continuities across performance in secondary school assessments, undergraduate medical school performance, and postgraduate examination performance in the form of MRCP(UK),³² with a preliminary analysis suggesting that MRCGP also correlates in a similar

way. This we have called the “academic backbone,” and it suggests that medical training is part of a continual acquisition of what we have called “medical capital.”

Better performance on the two parts of PLAB correlates with better performance on the various parts of MRCP(UK) and of MRCGP (table 3³). There is also specificity in that the knowledge based assessment of PLAB Part 1 particularly correlates with MRCP(UK) Part 1 and MRCGP AKT, whereas the clinical assessment of PLAB Part 2 correlates better with MRCP(UK) PACES and MRCGP CSA, both of which are clinical assessments. PLAB therefore has predictive validity for MRCP(UK) and MRCGP.

For comparative purposes, table 3³ also shows correlations between the separate parts of MRCP(UK) and MRCGP for those candidates who happen to have taken both assessments.³³ Again there is specificity, with knowledge based assessments correlating highly ($r=0.673$ between MRCP(UK) Part 1 and the AKT), and the clinical examinations (PACES and the CSA) also correlating highly ($r=0.496$). The latter correlation is particularly important, as it suggests that the modest correlation between PLAB Part 2 and both PACES ($r=0.186$) and the CSA ($r=0.321$) is not a reflection of poor correlation between clinical assessments in general but is more likely explained by the relatively low reliability of PLAB Part 2, which unpublished analyses suggest is in the range 0.55 to 0.71.

Outcome equivalence of MRCP(UK) and MRCGP candidates who are international medical graduates or UK medical graduates

If UK and PLAB graduates are outcome equivalent then the simplest of predictions is that their mean scores on the MRCP(UK) or MRCGP assessments should be the same. Table 4⁴ shows that they are not. For all of the assessments, the mean marks of PLAB graduates are substantially below those of UK graduates. The size of the effect is calculated as Glass's Δ (the difference in the mean scores divided by the standard deviation of the reference group, which here is the UK graduates). Glass's Δ is -0.94, -0.91, and -1.40 for MRCP(UK) Part 1, Part 2, and PACES, and -1.01 and -1.82 for MRCGP AKT and CSA. A conventional classification describes effect sizes of greater than 0.8 as “large,” and these values are undoubtedly substantial, the average Glass's Δ of -1.22 meaning there is about one and a quarter standard deviations between the UK and the PLAB groups.

The finding of a clear difference in performance of UK and PLAB candidates, coupled with a good correlation between PLAB scores and subsequent performances in MRCP(UK) and MRCGP, raises the immediate question of what a pass mark for PLAB might need to be, all other things being equal, to achieve outcome equivalence between UK and PLAB graduates. We therefore describe two separate methods of estimating a pass mark that would result in equivalent subsequent performance of UK and PLAB graduates on the MRCP(UK) and MRCGP examinations.

Method 1: Equating to median performance of UK graduates

A typical UK graduate taking MRCP(UK) Part 1 is at the median level of performance on that assessment, so that half of UK graduates perform better and half perform less well. Figure 2⁵ shows how the equivalent median for PLAB graduates may be estimated.

- The distribution of marks of UK graduates taking MRCP(UK) Part 1 is shown at the far right in blue.

- On a scale relative to the pass mark of zero, their median mark is +1.03, shown as the thick horizontal red line, so that UK graduates are therefore slightly more likely to pass than to fail MRCP(UK) Part 1 on their first attempt.
- The marks of the PLAB graduates at MRCP(UK) Part 1 are shown in the pale yellow histogram, third from the right.
- This distribution is clearly shifted downwards relative to the UK graduates, and the mark of +1.03, which is at the median for UK graduates, is on the 81st centile of the PLAB graduates.
- The horizontal orange histogram at the bottom shows the distribution of marks at first attempt on PLAB Part 1 by PLAB graduates.
- Finding a pass mark that results in equivalence with the UK distribution requires a pass mark to be set at PLAB Part 1, which results in a distribution of MRCP(UK) Part 1 scores in PLAB graduates which has a median of +1.03, the same as that for UK graduates.
- That can be estimated by considering only PLAB graduates with a mark higher than some threshold, which can be adjusted until the median of those taking MRCP(UK) Part 1 is +1.03.
- The dark green vertical line in fig 2 is set at such a threshold (“pass mark”) of +25.
- The MRCP(UK) Part 1 marks of all those PLAB graduates to the right of the dark green line are shown in the middle, pale green histogram at top right, and for this group the median is very close to +1.03, half being above that value and half below it.

On that basis, a pass mark for PLAB Part 1 of +25 compared with the present pass mark (which is defined as zero) would result in a group of PLAB graduates performing equivalently on MRCP(UK) Part 1 to UK graduates. Of the 7823 PLAB graduates taking MRCP(UK) Part 1, only 1409 (18.0%) are in the green distribution.

A similar analysis can be carried out for MRCP(UK) Part 2 in relation to PLAB Part 1. For UK graduates the median is +6.01, a value which is at the 82nd centile for PLAB graduates. Adjusting the threshold for PLAB Part 1 until the PLAB graduates have a median of +6.01 requires a threshold of +32 compared with the present PLAB Part 1 pass mark of zero; on that basis, 516 of the 5133 PLAB graduates currently taking MRCP(UK) Part 2 are equivalent to UK graduates (10.1%).

MRCP(UK) PACES is more problematic for calculating an equivalent threshold. The UK graduates have a median mark of +2.0 on PACES, a mark that is at the 91st centile for PLAB graduates. However it is not possible to get a threshold for PLAB Part 2 which produces a median of +2.0, there simply being no candidates left. The best that can be said therefore is that the threshold is $>+18$.

The analyses for MRCPGP are similar. The median AKT mark for MRCPGP UK graduates is 21 and for PLAB graduates is 5, and the median CSA mark for MRCPGP UK graduates is 14 and that for PLAB graduates is -5. To achieve an equivalently performing median candidate as between UK graduates and PLAB candidates on first attempt would require the pass mark for PLAB Part 1 to be increased by +35 marks and that for PLAB Part 2 to be increased by +10 marks. Using these values as a pass mark would result in many fewer PLAB graduates taking MRCPGP, 106 of the 3160 taking AKT (3.4%) and 114 of the 1388 taking CSA (8.2%).

Method 2: Comparison with performance of graduates from different UK medical schools

The second method takes a rather different approach. In a previous analysis of the performance of graduates of different UK medical schools at MRCP(UK)²⁴ there were clear and large differences in performance at MRCP(UK) between graduates of different medical schools. That result extended and developed the much earlier analysis of Wakeford et al³⁴ for MRCPGP, and has been repeated in recent analyses of the MRCPGP.³⁵ Similar differences between medical schools have also been reported for FRCA³⁶ and MRCOG.¹⁶ The ordering across medical schools is broadly similar in all of the studies, with some variation due to sampling differences, and perhaps also differences in medical school training.

Our second method addressed the question of equivalence by estimating the level of performance at PLAB which results in a similar performance to that of graduates from the various UK medical schools. The PLAB graduates have therefore been divided into 12 equally spaced subgroups according to performance at PLAB Part 1 (or Part 2), which groups can then be compared with graduates of individual UK medical schools. Subgroups were based on steps of five marks for PLAB Part 1 and steps of three marks for PLAB Part 2, so that groups can be directly compared with the marking scales for each assessment.

Figure 3 shows results for MRCP(UK) Part 1. The blue points show performance of graduates of UK medical schools, ranked from highest to lowest. New medical schools, whose graduates have not been taking MRCP(UK) for long enough to establish stable patterns in their results, have been omitted as numbers are not yet large enough to have reasonable standard errors. Differences between UK medical schools are highly significant, as can be seen from the narrowness of the 95% confidence intervals, but they are not of direct interest here. PLAB graduates are shown as the 12 red points, corresponding to the different grouping of marks at the first attempt at PLAB Part 1, relative to the pass mark. A separate group of EEA graduates who are not required to take PLAB is shown as a green point. PLAB groups were regarded as equivalent to UK medical schools if, using the Ryan-Einot-Gabriel-Welsch Q post hoc test, their performance was not significantly different from a UK medical school.

The highest scoring PLAB group (“PLAB1 A1 35+” which scored ≥ 35 marks above the PLAB1 pass mark) has a mean performance equivalent or better than the mean performance of graduates of all but two of the UK medical schools (Oxford and Cambridge) and is clearly achieving very highly. Similarly the second and third groups (“PLAB1 A1 30-34” and “PLAB1 A1 25-29”) have a mean performance better than or similar to the graduates of many UK medical schools. The fourth group (“PLAB1 A1 20-24”) with PLAB scores 20-24 points above the pass mark, has a mean performance that is not distinguishable from the mean performance of the lowest performing UK medical school, using the Ryan-Einot-Gabriel-Welsch Q post hoc test. The eight remaining PLAB1 groups, from “PLAB1 A1 15-19” downwards, all perform at a significantly lower average level than graduates of any of the UK medical schools. Taken overall, figure 3 suggests that the top four PLAB1 groups are equivalent to graduates from UK medical schools when taking MRCP(UK) Part 1, whereas the lower eight groups perform less well. Those results suggest that an equivalence level is +20 to +24 points above the current pass mark. In this figure (and figs 4-7) an orange dotted line marks the mean scores of the lowest performing UK university or medical school.

Similar calculations can be carried out for MRCP(UK) Part 2 in relation to PLAB Part 1, and for MRCP(UK) PACES in relation to PLAB Part 2, and plots are shown in figures 4 and 5. For MRCP(UK) Part 2 (fig 4), the mean performance of only the top two PLAB groups (30-34 and 35+) is equivalent to that of UK graduates, making +30 to +34 the likely equivalence. For PACES (fig 5), only the top two groups are equivalent to graduates of UK medical schools, making the equivalence +16 to +18.

Analyses for the MRCGP are shown in figures 6 and 7, comparing performance in the AKT in relation to PLAB Part 1 and in the CSA in relation to PLAB Part 2. Note that the MRCGP database subdivides London medical schools.

For the AKT (fig 6), the top PLAB group, with PLAB Part 1 scores of ≥ 35 above the pass mark, is clearly equivalent to many UK medical schools, as are all of the top five groups including PLAB Part 1 mark 15-19. However the group with PLAB Part 1 mark 10-14 is performing significantly less well. A probable equivalence is therefore at +15 to +19.

For the CSA (fig 7), only the PLAB Part 2 16-18 group and the (very small) PLAB Part 2 18+ group are equivalent to the lowest scoring UK medical school, suggesting that PLAB Part 2 scores of +16 to +18 would be necessary for equivalence.

Summary: overall estimate of PLAB1 and PLAB2 pass marks for outcome equivalence

A simple comparison of the mean performance of UK and PLAB graduates on MRCP(UK) and MRCGP makes clear that there is not outcome equivalence, PLAB graduates perform less well by about one and a quarter standard deviations. Two methods are described for estimating how PLAB pass marks could be altered to result in outcome equivalence, both making the assumption that all other factors are similar between the two groups. The method of equating medians, and comparing performance with graduates of individual UK medical schools, give slightly different results, which are summarised in table 5.

Estimating an equivalence level of PLAB Part 1 for the knowledge assessments of MRCP(UK) Parts 1 and 2 and MRCGP AKT using the two methods suggests overall that a pass mark of the order of +27 marks higher than at present would result in outcome equivalence (+31 based on method 1 and +24 based on method 2). Since the PLAB Part 1 typically has nearly 200 questions, in terms of percentage of items correct, the pass mark would need to be moved from its present level of about 63% to about 76%.

For PLAB Part 2, both methods find that a considerably higher PLAB pass mark would be needed to achieve outcome equivalence, there being barely any level of attainment at PLAB Part 2 which is equivalent to the performance of UK graduates, so that only the very top performers seem to be equivalent to UK graduates. Averaging across the estimates, the pass mark would seem to need to rise by about +15 to +16 marks (+14 based on method 1 and +17 based on method 2). Some of the problems with estimating may result either from the assessment not stretching candidates at the top end, or from the relatively low reliability of Part 2, an aspect of the assessment which inevitably makes its predictive power less than is desirable.

It should be reiterated that these calculations make the assumption that the only differences between the groups are in the pass mark for PLAB (see discussion).

The role of IELTS on performance in PLAB and in MRCP(UK) and MRCGP

PLAB candidates have mostly attained the required level at IELTS, although some are exempted. Since PLAB is an assessment carried out in English, as are MRCP(UK) and MRCGP, an important question concerns the extent to which poor performance at later postgraduate qualifications may be mediated via problems with English. We have investigated that for both MRCP(UK) and MRCGP, but only report here the results for MRCGP. Few PLAB candidates had IELTS scores below 7 or over 8, and we therefore divided the candidates into three groups: ≤ 7 , 7.5, and ≥ 8 .

Figure 8 shows performance at the MRCGP AKT in relation to performance at PLAB Part 1 at the first attempt and the IELTS level, the "traffic lights" showing that, at most levels of PLAB 1 performance, those with the highest IELTS scores (green) perform better than those with the lowest IELTS scores (red). IELTS is clearly therefore important. However a multiple regression shows that the predictive effect of PLAB Part 1 ($\beta=0.496$) is very much stronger than the effect of IELTS ($\beta=0.086$).

Figure 9 shows a similar analysis for performance at the MRCGP CSA, broken down by PLAB Part 2 performance at first attempt and IELTS level, shown as traffic lights. The effects are somewhat less clear, in some cases due to smaller sample sizes. Again, the multiple regression shows the effect of PLAB Part 2 ($\beta=0.278$) is stronger than that for IELTS ($\beta=0.187$). The lower effect of PLAB Part 2 (compared with PLAB Part 1 on the AKT) probably reflects the lower reliability of PLAB Part 2, and the larger effect of IELTS is probably due to the greater importance of language, particularly spoken language, in a clinical examination.

Discussion

The results of this data linkage study show that there are good correlations between PLAB and the subsequent assessments of MRCP(UK) and MRCGP, which means that PLAB is a valid assessment of skills relevant to progression during UK postgraduate training (that is, there is construct equivalence). It is also clear that, compared with UK graduates, PLAB graduates perform less well in two major postgraduate medical examinations in the UK, so that there is not criterion related outcome equivalence. Outcome equivalence could be produced were the pass mark for PLAB to be set at a higher level than it currently is, although that and the other conclusions have important caveats and need to be interpreted with care.

Construct equivalence

Performance in PLAB Part 1 correlates well with subsequent performance on MRCP(UK) and MRCGP, suggesting that the constructs it is measuring are parallel to those assessed by the two postgraduate examinations, and MRCP(UK) and MRCGP also correlate strongly in candidates who take both assessments.³³ Additionally, as table 3 shows, knowledge tests shows stronger correlations with other knowledge tests, and clinical tests with other clinical tests, suggesting that knowledge and clinical tests assess separate but related domains. The three assessments are therefore measuring similar underlying constructs of knowledge and clinical skills. PLAB is reliably putting candidates into a meaningful order that predicts other postgraduate outcomes in a way which is probably similar to UK medical school finals.³² PLAB therefore seems to act in a similar way to the "academic backbone" which underpins secondary school, undergraduate, and postgraduate performance in UK medical students and

graduates,³² and which involves the continual development of the skills, knowledge, and expertise that underpin competent medical practice—the acquisition of “medical capital.”

Criterion related outcome equivalence

That PLAB graduates do not progress through their careers in the same way as UK graduates seems clear from table 4. That the difference is not merely in the summative assessments for these two specialties is shown by the parallel analysis of PLAB in relation to the Annual Review of Competence Progression,²⁵ which suggests that the lack of progression is in workplace based assessments and deanery assessments of progress across the entire range of medical specialties. PLAB graduates do not therefore show criterion related outcome equivalence.

Explaining the lack of outcome equivalence is more complex, and several factors are considered below.

The extent of non-equivalence

The extent of non-equivalence can be evaluated numerically by considering at what level the PLAB pass mark would need to be set in order to produce outcome equivalence. We have described two methods. One of our methods considers the performance of a median UK graduate at MRCP(UK) and MRCGP and asks what PLAB pass mark would result in a median PLAB graduate performing at the same level in MRCP(UK) and MRCGP as that median UK graduate (see table 5).

A potential problem with such a method is that it could be argued that equivalence should be set not at the median UK graduate but at some lower value, such as, say, the fifth centile of ability of UK graduates. Considering the right hand part of figure 2, the fifth centile for UK graduates taking MRCP(UK) Part 1 was a mark of -16.2 , only 5% of UK graduates scoring less than that. However, of the PLAB graduates, 26.1%, over five times as many, scored below -16.2 . Using the method described previously, a threshold at PLAB Part 1 could also be found at which 95% of PLAB graduates score -16.2 or above. When that is done, the threshold is $+27$, a value slightly higher than the $+25$ we reported in table 5 for the median. The approach could be extended so that PLAB graduates were required to be equivalent to the worst performing UK graduate, but any such analysis of extreme values would be vulnerable to random sampling variation. Overall, the similarity of estimates based on the median and the fifth centile suggests that equivalence calculated on other centiles would have similar results to those presented here.

Although the median equivalence method we have described seems to be robust for PLAB Part 1, we note that both PLAB Part 1 and MRCP(UK) have reliabilities that are above 0.9.³⁰ PLAB Part 2 has a rather lower reliability, at about 0.55 to 0.71, and that clearly has consequences for the calculation of equivalence. In an extreme case, were PLAB Part 2 to have a reliability of zero then no threshold could ever result in equivalence as all subsamples above any threshold would necessarily have the same mean. Further consideration and modelling of the effects of reliability on estimating equivalence is therefore desirable.

Our second method of evaluating non-equivalence takes known differences between the graduates of different UK medical schools as its starting point, some medical schools having graduates who consistently perform better at postgraduate medical examinations than do others, with about two thirds of the variance of those differences probably being due to differences in qualifications at entry to medical school.²⁴ No

doubt, were it possible to estimate similar findings for the medical schools attended worldwide by international medical graduates, then similar differences would probably be found. International medical graduates taking PLAB are not, though, either from a random sample of international medical schools nor are they a random sample of graduates from those medical schools. International medical graduates wish to practise in the UK for a host of different reasons.

Our division of PLAB graduates into 12 groups based on Part 1 and Part 2 performance allows comparison with performance of graduates from UK medical schools. Our equivalence criterion of not having a mean performance significantly lower than any UK medical school is a first attempt at using such a method, and, although there may be an argument that it is unreasonably conservative, it is also the case that all of the UK medical schools have been inspected by the GMC and their graduates found to have acceptable performance standards, whereas foreign medical schools are not subject to that inspection. As with our first method, our second method of evaluating non-equivalence could probably be carried out with many variations on the basic theme, and that requires future exploration.

Factors potentially influencing outcome non-equivalence

Although PLAB graduates and UK graduates do not show outcome equivalence, interpreting and explaining that difference is not straightforward, and a number of possible moderating factors need to be considered. The calculations of our two methods assess at what level the PLAB pass marks might need to be set in order to produce outcome equivalence. The calculations make a number of assumptions, and care must be taken in interpreting their numerical estimates. The most crucial assumption, as ever, is “all other things being equal.” However, all other things cannot be assumed to be equal, although the extent of the inequality is not known precisely. The numbers in table 5 should therefore be considered as upper limits of where the PLAB pass marks may need to be set. Factors that need to be considered in interpreting the results include the following.

Demographic differences

PLAB graduates inevitably differ demographically from UK graduates in many ways (see table 2), and some of those ways may correlate with performance in PLAB and in subsequent assessments. Although such demographic differences are not to be disputed, they are mostly not relevant to the primary topic of this report, which is the assessment of outcome equivalence. The stated role of PLAB is to allow only doctors who are equivalent to UK graduates to enter UK medical training and practise, and if there is equivalence then progression of PLAB graduates should also be equivalent to that of UK graduates. The influence of demographic factors may be of sociological interest for understanding and explaining differences, but the purpose of postgraduate examinations is to maintain absolute standards for all doctors, which necessarily will be irrespective of demography. Unless it is deemed appropriate that professional standards are to be set at different levels for different demographic groups, then demographic variables should not be taken into account in the statistical analyses.

English language proficiency

The PLAB examinations, as well as MRCP(UK) and MRCGP, are examinations taken in English, and inevitably it is a concern that doctors with high levels of clinical competence might be

being excluded because of language problems. However, as Esmail and Roberts wrote of the CSA, “it is designed to ensure that doctors are safe to practise in [the] UK,” and PLAB is also designed with a similar objective, English being the language in which most consultations and professional interactions take place in the UK. Language ability, as assessed by IELTS, does have some influence on MRCP(UK) and MRCGP outcome, but the effects are small, and overall the conclusion is probably similar that of the 1986 PLAB review: “The failure of candidates was due in the main part to their lack of professional knowledge rather than difficulty in communicating in English.”²²

Postgraduate training and experience

All doctors taking the MRCGP have taken part in an approved, three year deanery training programme which takes places after foundation year 2, with the AKT exam taken after two years and CSA after three years. In contrast, MRCP(UK) is not restricted to doctors on training programmes, although many candidates are in core medical training programmes. Performance differences in MRCP(UK) and MRCGP may in part reflect differences in the quality of international medical graduate and UK medical graduate postgraduate training programmes. Deaneries undoubtedly differ in the proportion of international medical graduates on their general practice training programmes, and there are also differences in success rates.⁶ Training schemes within deaneries probably also vary in quality, and it is possible that international medical graduates are allocated to poorer quality training (and one of us elsewhere has referred to, “the inverse care law of training ... in which those who most need the added value of education are assigned to the least popular schemes”³⁷). The quality of postgraduate education cannot straightforwardly be taken into account without direct measures of the quality of training programmes and schemes (and the GMC’s National Training Survey might in principle provide such measures, particularly if linked to examination databases). It might also be the case that differences in postgraduate outcome between UK medical schools are in part due to differences in training programme quality.³⁸ Clearly there is an urgent need to take training posts into account.

Direct methods for estimating entry equivalence

Standard setting is an imperfect science. Our analysis of outcomes in relation to PLAB attainment levels was motivated by Kane’s “direct, criterion-related approach” for standard setting, and it was used because no direct method of assessing entry equivalence is currently available in the UK. The most direct method of assessing entry equivalence would be if the UK had a national qualifying examination that was also taken by international medical graduates, unchanged and with the same pass mark as that for UK graduates.

The standard for PLAB is currently set by the Angoff method for PLAB Part 1, and by a borderline group method for Part 2. Both methods are well recognised in the literature,^{39 40} but Angoff in particular has potential problems.^{27 41-43} A standard setting method can be valid, but that does not ensure that an implementation of the method is valid or appropriate. Using Kane’s terminology,²⁶ there may be acceptable procedural validity and internal validity, but they cannot guarantee that a pass mark is set at the right level.²⁷ Ultimately the validity of a pass standard is an empirical matter to be assessed by its relationship to standards set by other methods for other parallel assessments.

Of its very nature, PLAB is an assessment similar to those carried out in medical schools throughout the UK. Just as examining boards at secondary school level work closely together to ensure that standards on assessments such as A levels are comparable, using a mixture of statistical and evaluative methods,²³ so PLAB and other equivalent qualifications such as medical school finals could collaborate on shared standard setting. A range of direct methods for equating standards is available, some of which rely on item overlap of assessments²² and of examiners. The inclusion of items from PLAB in medical school finals and vice-versa would help in the equating process.

Without any direct method of assessing entry equivalence, the only conclusion can be that there are no strong reasons to believe that PLAB standards are at the same level as foundation year 1, and the lack of outcome equivalence, with its large effect size, is compatible with the PLAB standard being set too low, although precisely by how much is difficult to assess accurately.

The relationship between entry equivalence and outcome equivalence

An implicit assumption in assessments of entry equivalence is that entry equivalence and outcome equivalence are directly related, the former ensuring the latter. Thus it is sometimes argued, for instance, that since all general practice trainees are under the supervision of UK postgraduate deaneries, and all those trainees have passed the PLAB exams (and hence there is entry equivalence with UK graduates), then PLAB graduates and UK graduates should also show outcome equivalence when taking MRCGP. Even were it the case that the standard of PLAB and foundation year 1 showed exact entry equivalence, that still would not ensure that, say, international medical graduates and UK medical graduates would show outcome equivalence. As a concrete example, international medical graduates applying to and selected by one deanery, who had entry equivalence to UK medical graduates, having passed the same selection tests with the identical pass mark, had lower mean levels of attainment on the selection tests³⁷, so that outcome equivalence could not be expected. Entry equivalence can only ensure outcome equivalence in different groups if the distribution of marks in those groups is the same, and when distributions are not the same then outcome equivalence will not occur despite entry equivalence.

Strengths and weaknesses of this study

The strength of the present study is that it looks at the marks of a large number of international medical graduates who have taken PLAB and compares the marks of both international medical graduates and UK medical graduates on two major postgraduate assessments, which together are taken by over half of the UK medical workforce. The data linkage allows generalisable insights into PLAB that were hitherto unavailable, and in particular it suggests that the pass mark may not be appropriate, although the validity of PLAB is affirmed. A potential weakness of the present study is that it includes data from only two Royal College examinations, but the separate analysis of Annual Review of Competence Progression data, which includes doctors from all specialties and includes non-examination outcomes, supports the present findings.²⁵ A weakness of the present study is that it has no information about the training programmes and schemes on which UK medical graduates and international medical graduates have been based, and if international medical graduates systematically have lower quality training, then that may account for some of the effects reported here.

A narrative interpretation

The present study has raised many issues concerning international medical graduates and their selection and training. The following paragraph provides a synoptic overview and an interpretation of the various issues.

International medical graduates undoubtedly perform less well at MRCP(UK), MRCGP, and Annual Review of Competence Progression, and probably at other postgraduate examinations. That seems unlikely to result from systematic examiner bias or discrimination, not least as the effect size is large, being over one standard deviation. Some of the difference may well be due to differences in training programmes, with international medical graduates systematically being allocated to less good training programmes due to inequitable access. However, training programmes would have to be extremely disparate in their effects to produce an effect size of over one standard deviation. Other factors, such as language ability, may correlate with outcome, but probably also correlate to a large extent with prior medical knowledge, and anyway should in large part have been taken into account by PLAB and are legitimate reasons for examination failure. The PLAB pass mark is intended to be set at the same level as foundation year 1, but there are no formal mechanisms to ensure that beyond the judgments of a standard setter. Standard setter judgments have no formal mechanism for aligning or comparing them with assessments such as medical school finals, such as item sharing or examiner sharing. It is therefore plausible that the PLAB pass mark is set too low, there being little evidence to justify its current level. Even if there were strict entry equivalence of PLAB and foundation year 1, the distributions of those taking the assessments are almost certainly different, with only a small proportion of medical students failing finals compared with a much higher proportion of international medical graduates taking PLAB Part 1 and 2. Without similar distributions, and even with entry equivalence, and even if training and all other factors were the same, international medical graduates would still be expected to perform less well at outcome.

Conclusions

PLAB and its predecessor, TRAB, throughout their 40 year history, have meant to be set at a standard equivalent to that of UK graduates, currently as at the end of foundation year 1. Although some early attempts were made at assessing equivalence by administering the test to UK medical students and doctors,² there have been no serious recent attempts at empirical assessment. Large scale record linkage has now allowed the sorts of comparison that are described here, and those data suggest that the standard for PLAB has in recent years been set too low if equivalent progression by PLAB graduates to UK graduates is expected and required. The standard for PLAB therefore needs reconsideration.

We cannot finish without acknowledging the various implications of these findings. The only concern of the Professional and Linguistic Assessments Board is with ensuring that the level of competence of international medical graduates is sufficient to ensure patient safety. PLAB graduates, though, currently form a sizeable proportion of the doctors entering the NHS, and any change in their numbers would inevitably have consequences for service delivery. Those implications cannot be a part of this study, but we acknowledge that they are potentially problematic. Nevertheless, getting the standard of PLAB at a correct level is fundamental to ensuring the quality of postgraduate medical education and training, the delivery of

medical care of the highest quality, and thus ensuring patient safety in the NHS.

We thank Michael Harriman, Katharine Lang and William Curnow for their help in understanding the processes of the PLAB examinations; Daniel Smith and Andy Knapton for help with data merging, and Daniel Smith for additional data analyses; Thomas Jones for facilitating various aspects of the study; Jane Dacre and the Royal Colleges of Physicians for giving permission for the analysis of MRCP(UK) data to take place, and Liliانا Chis for helping in the preparing of those data; and Sue Rendel and the Royal College of General Practitioners for their permission to analyse the MRCGP data. We particularly thank Daniel Smith for his careful reading of the manuscripts.

Contributors: ICM is the guarantor for the study. The original idea for the study came through discussions between ICM and RW. ICM was responsible for the linkage of the MRCP(UK) and PLAB data, and RW was responsible for the linkage of the MRCGP and PLAB data. ICM and RW worked together on their analyses. The current paper is based on the report which ICM and RW submitted to the PLAB Working Party. The manuscript has been produced collaboratively by the authors. Collectively ICM, RW, and the General Medical Council are responsible for the data, although not all have seen all data.

The report on which this paper was based has been seen by members of General Medical Council staff and members of the PLAB Working Party, and useful comments received from them.

Funding: No funding was made available to ICM and RW for the present study, but ICM has received attendance allowances for meetings of the PLAB Working Party.

Competing interests: Both authors have completed the Unified Competing Interest form at ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: ICM is a member of the General Medical Council's Working Party on the PLAB assessment and has received attendance fees from the GMC, and is educational advisor to the MRCP(UK); RW has been an educational advisor to the MRCGP since 1984.

Ethical approval: Not required..

Data sharing: No additional data are available.

Transparency: The lead author, the manuscript's guarantor, affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that there are no discrepancies from the study as planned.

- 1 IELTS. International English Language Testing System. 2013. www.ielts.org/.
- 2 General Medical Council. Report of the working party on the PLAB tests [January 1986; chair Sir David Innes Williams]. GMC, 1986.
- 3 General Medical Council. Steering group on the review of the PLAB test. Report: November 1999. GMC, 1999.
- 4 PLAB Test Review Group. Final report [July 2004; chair Lord Patel]. General Medical Council, 2004.
- 5 General Medical Council. Broad themes of the review and members of the working group. GMC, 2013. www.gmc-uk.org/static/documents/content/Broad_themes_of_the_review_and_members_of_the_working_group.pdf.
- 6 Esmail A, Roberts C. Independent review of the Membership of the Royal College of General Practitioners (MRCGP) examination. General Medical Council, 2013. www.gmc-uk.org/MRCGP_Final_Report_18th_September_2013.pdf. 53516840.pdf.
- 7 Esmail A, Roberts C. Academic performance of ethnic minority candidates and discrimination in the MRCGP examinations between 2010 and 2012: analysis of data. *BMJ* 2013;347:f5662.
- 8 Woolf K, Potts HWW, McManus IC. The relationship between ethnicity and academic performance in UK-trained doctors and medical students: a systematic review and meta-analysis. *BMJ* 2011;342:d901.
- 9 Woolf K, McManus IC, Potts HWW, Dacre JE. The mediators of minority ethnic underperformance in final medical school examinations: a longitudinal study. *Br J Educ Psychol* (forthcoming).
- 10 Woolf K, Potts HWW, Patel S, McManus IC. The hidden medical school: a longitudinal study of how social networks form, and how they relate to academic performance. *Med Teacher* 2012;34:577-86.
- 11 Denney ML, Freeman A, Wakeford R. MRCGP Clinical Skills Assessment: are the examiners biased, favouring their own by sex, ethnicity and degree source? *Br J Gen Pract* 2013; doi:10.3399/bjgp13X674396.
- 12 Dewhurst NG, McManus IC, Mollon J, Dacre JE, Vale JA. Performance in the MRCP(UK) examination 2003-4: analysis of pass rates of UK graduates in the clinical examination in relation to self-reported ethnicity and gender. *BMC Med* 2007;5:8. doi:10.1186/1741-7015-5-8.

What is already known on this topic

International medical graduates perform less well than UK graduates on a number of postgraduate examinations in the UK, but the reasons for this are not clear

International medical graduates who wish to practise in the UK can enter the Medical Register by passing the PLAB assessments, which the General Medical Council expects to be at an equivalent level to that of UK graduates at the end of the first year of foundation training

The General Medical Council has asked whether the subsequent career progression of international medical graduates passing PLAB is equivalent to that of UK graduates

What this study adds

A data linkage study allows performance at PLAB to be related to performance in the MRCP(UK) and MRCGP examinations and showed that PLAB performance is a good predictor of subsequent MRCP(UK) and MRCGP performance

However, PLAB graduates perform substantially less well at MRCP(UK) and MRCGP than UK graduates, with an effect size of over one standard deviation, so that career progression is not equivalent

Career progression in terms of examination performance could be made equivalent were the pass mark for PLAB to be raised considerably, but this would though produce a fall in pass rates with implications for the health service workforce

- 13 McManus IC, Elder AT, Dacre J. Investigating possible ethnicity and sex bias in clinical examiners: an analysis of data from the MRCP(UK) PACES and nPACES examinations. *BMC Med Educ* (forthcoming).
- 14 MRCP(UK). MRCP(UK) annual review 2011/2012. Federation of Royal Colleges of Physicians, 2013. [www.mrcpuk.org/SiteCollectionDocuments/MRCP\(UK\)%20Annual%20Review%202011.pdf](http://www.mrcpuk.org/SiteCollectionDocuments/MRCP(UK)%20Annual%20Review%202011.pdf).
- 15 Tyrer SP, Leung W-C, Smalss J, Katona C. The relationship between medical school of training, age, gender and success in the MRCPsych examinations. *Psychiatric Bull* 2002;26:257-63.
- 16 Rushd S, Landau AB, Khan JA, Allgar V, Lindow SW. An analysis of the performance of UK medical graduates in the MRCOG Part 1 and Part 2 written examinations. *Postgrad Med J* 2012;88:249-54.
- 17 Rushd S, Landau AB, Lindow SW. An evaluation of the first time performance of international medical graduates in the MRCOG Part 1 and Part 2 written examinations. *Eur J Obstet Gynecol Reprod Biol* 2013;166:126.
- 18 General Medical Council. Annual specialty report for 2010/11 [Royal College of Paediatrics and Child Health]. GMC, 2011. www.gmc-uk.org/RCPC_H_ASR.pdf_48272553.pdf.
- 19 General Medical Council. Review of the routes of the GP and specialist register: final report. GMC, 2012.
- 20 Rogers JL, Howard KI, Vessey JT. Using significance tests to evaluate equivalence between two experimental groups. *Psychol Bull* 1993;113:553-65.
- 21 Cleland J, Dowell J, McLachlan J, Nicholson S, Patterson F. Identifying best practice in the selection of medical students (literature review and interview survey). General Medical Council, 2012. www.gmc-uk.org/Identifying_best_practice_in_the_selection_of_medical_students.pdf_51119804.pdf.
- 22 Kolen MJ, Brennan RL. *Test equating: methods and practices*. Springer-Verlag; 1995.
- 23 Newton P, Baird J-A, Goldstein H, Patrick H, Tymms P. Techniques for monitoring the comparability of examination standards. Qualifications and Curriculum Authority, 2007.
- 24 McManus IC, Elder AT, De Champlain A, Dacre JE, Mollon J, Chis L. Graduates of different UK medical schools show substantial differences in performance on MRCP(UK) Part 1, Part 2 and PACES examinations. *BMC Med* 2008;6:5.
- 25 Tiffin PA, Illing J, Kasim AS, McLachlan JC. Annual Review of Competence Progression (ARCP) performance of doctors who passed Professional and Linguistic Assessments Board (PLAB) tests compared with UK medical graduates: national data linkage study. *BMJ* 2014;348:g2622.
- 26 Kane M. Validating the performance standards associated with passing scores. *Rev Educ Res* 1994;64:425-61.
- 27 Verheggen MM, Muijters AMM, Van Os J, Schuwirth LWT. Is an Angoff standard an indication of minimal competence of examinees or of judges? *Adv Health Sci Educ* 2008;13:211.
- 28 PACES: Practical Assessment of Clinical Examination Skills. The new MRCP(UK) clinical examination. *J R Coll Physicians Lond* 2000;34:57-60.
- 29 Elder A, McAlpine L, Bateman N, Dacre J, Kopelman P, McManus IC. Changing PACES: developments to the examination in 2009. *Clin Med* 2011;11:231-4.
- 30 Tighe J, McManus IC, Dewhurst NG, Chis L, Mucklow J. The standard error of measurement is a more appropriate measure of quality in postgraduate medical assessments than is reliability: An analysis of MRCP(UK) written examinations. *BMC Med Educ* 2010;10:40.
- 31 McManus IC, Ludka K. Resitting a high-stakes postgraduate medical examination on multiple occasions: nonlinear multilevel modelling of performance in the MRCP(UK) examinations. *BMC Med* 2012;10:60.
- 32 McManus IC, Woolf K, Dacre J, Paice E, Dewberry C. The academic backbone: longitudinal continuities in educational achievement from secondary school and medical school to MRCP(UK) and the specialist register in UK medical students and doctors. *BMC Med* 2013;11:242.
- 33 Wakeford R, Denney ML, Ludka-Stempien K, Dacre J, McManus IC. Cross-validation of MRCGP and MRCP(UK): a database linkage study of 2,284 candidates taking both assessments. *BMC Med Educ* (submitted).
- 34 Wakeford R, Foulkes J, McManus IC, Southgate L. MRCGP pass rate by medical school and region of postgraduate training. *BMJ* 1993;307:542-3.
- 35 Royal College of General Practitioners. MRCGP annual reports. RCGP, 2013. www.rcgp.org.uk/gp-training-and-exams/mrcgp-exam-overview/mrcgp-annual-reports.aspx.
- 36 Bowhay AR, Watmough SD. An evaluation of the performance in the UK Royal College of Anaesthetists primary examination by UK medical school and gender. *BMC Med Educ* 2009;9(38).
- 37 Wakeford RE. International medical graduates' relative under-performance in the MRCGP AKT and CSA examinations. *Educ Primary Care* 2012;23:148-52.
- 38 Wembridge HA. Experiment and statistics in the selection of employees: methods used in the selection of machine operators in a clothing plant. *J Am Stat Assoc* 1923;18:600-6.
- 39 Downing SM, Tekian A, Yudkowsky R. Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. *Teaching Learning Med* 2006;18:50-7.
- 40 Ricker KL. Setting cut-scores: a critical review of the Angoff and modified Angoff methods. *Alberta J Educ Res* 2013;52:53-64.
- 41 Clauser BE, Mee J, Baldwin SG, Margolis MJ, Dillon GF. Judges' use of examinee performance data in an Angoff standard-setting exercise for a medical licensing examination: an experimental study. *J Educ Measure* 2009;46:390-407.
- 42 Mee J, Clauser BE, Margolis MJ. The impact of process instructions on judges' use of examinee performance data in Angoff standard setting exercises. *Educ Measure Issues Pract* 2013;32:27-35.
- 43 Margolis MJ, Clauser BE. The impact of examinee performance information on judges' cut scores in modified Angoff standard-setting exercises. *Educ Measure Issues Pract* 2014;33:15-22.

Accepted: 04 April 2014

Cite this as: *BMJ* 2014;348:g2621

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/>.

Tables

Table 1 | Comparison of performance of PLAB graduates known to have taken MRCP(UK) or MRCGP, or both, with PLAB graduates not known to have taken either examination within the time windows*

	Mean (SD) score on first attempt at PLAB relative to pass mark	No of graduates
PLAB Part 1		
Neither MRCP(UK) or MRCGP taken	9.19 (18.16)	15 323
MRCGP taken	7.01 (18.57)	1761
MRCP(UK) taken	10.87 (17.43)	6533
MRCP(UK) and MRCGP taken	7.53 (17.42)	1234
All PLAB graduates	9.39 (18.00)	24 851
PLAB Part 2		
Neither MRCP(UK) or MRCGP taken	6.33 (4.57)	15 323
MRCGP taken	7.25 (4.35)	1761
MRCP(UK) taken	6.13 (4.56)	6533
MRCP(UK) and MRCGP taken	6.49 (4.48)	1234
All PLAB graduates	6.35 (4.56)	24 851

*As the sampling periods for the PLAB and college exams differed, the distribution of the dates the PLAB exams were passed was examined against whether the college exams were attempted. The table excludes cases with a PLAB Part 1 test date after 13 July 2006 (n=6171) and PLAB Part 2 test date after 12 June 2007 (n=6172) as the distributions suggest that some of the doctors may not have had a chance to take the GP exams having only recently been able to register and so gain a place on a GP training programme, which is a prerequisite for taking the exams.

Table 2 | Demographics of candidates taking MRCP(UK) and MRCGP exams

Variable	UK graduates		PLAB graduates		Significance
	Value	No of graduates	Value	No of graduates	
Mean (SD) age at qualification (years)	24.9 (2.2)	24 640	24.9 (2.0)	9798	NS
MRCP(UK) candidates					
Mean (SD) age at 1st attempt of exam (years):					
PLAB Part 1	N/A	—	28.7 (4.3)	9812	-
PLAB Part 2	N/A	—	29.4 (4.4)	9344	-
MRCP(UK) Part 1	26.9 (2.4)	18 532	30.3 (4.5)	7823	P<0.001
MRCP(UK) Part 2	27.7 (2.3)	14 094	31.5 (4.2)	5133	P<0.001
MRCP(UK) PACES	28.4 (2.3)	14 409	32.6 (4.1)	4388	P<0.001
Mean (SD) interval between 1 st attempts of exams (weeks):					
MRCP(UK) Part 1 and Part 2	50.0 (32.2)	12 091	101.3 (75.0)	3947	P<0.001
MRCP(UK) Part 2 and PACES	39.4 (23.5)	12 051	66.7 (49.7)	4138	P<0.001
% female	56.3	24 634	32.0	9802	P<0.001
% non-white ethnicity*	41.1	24 641	96.3	9804	P<0.001
% UK nationals	N/A	—	7.8	9589	—
MRCGP candidates					
Mean (SD) age at 1st attempt of exam (years):					
PLAB Part 1	N/A	—	28.3 (4.4)	3067	-
PLAB Part 2	N/A	—	29.1 (4.5)	3067	-
AKT	30.5 (4.7)	10 044	34.4 (4.6)	3067	P<0.001
CSA	30.9 (4.2)	4481	36.1 (4.5)	1478	P<0.001
% non-white ethnicity*:					
AKT database	32.3	12 152	94.2	3160	P<0.001
CSA database	33.1	5924	94.4	1381	P<0.001
% female:					
AKT database	61.3%	10 048	44.8%	3067	P<0.001
CSA database	62.5%	4481	43.7%	1478	p<0.001
% UK nationals	N/A	N/A	12.0%	3233	—

*Self reported ethnicity from college databases, hence variation in numbers.

Table 3| Correlations of performance in PLAB Parts 1 and 2 with performance at MRCP(UK) and MRCGP and correlations between MRCP(UK) and MRCGP components. All assessments are at the first attempt, and all correlations are $P < 0.001$. Examinations are divided into knowledge assessments and clinical assessments

	PLAB Part 1 (knowledge)	PLAB Part 2 (clinical)	MRCGP AKT (knowledge)	MRCGP CSA (clinical)
MRCP(UK)				
Part 1 (knowledge)	$r=0.521$ (n=7823)	$r=0.194$ (n=7671)	$r=0.673^*$ (n=1988)	$r=0.348^*$ (n=1988)
Part 2 (knowledge)	$r=0.390$ (n=5133)	$r=0.227$ (n=4916)	$r=0.600^*$ (n=1131)	$r=0.386^*$ (n=1131)
PACES (clinical)	$r=0.171$ (n=4386)	$r=0.274$ (n=4120)	$r=0.471^*$ (n=943)	$r=0.496^*$ (n=943)
MRCGP				
AKT (knowledge)	$r=0.490$ (n=3160)	$r=0.186$ (n=3067)	N/A	N/A
CSA (clinical)	$r=0.232$ (n=1411)	$r=0.321$ (n=1388)	N/A	N/A

N/A=Not applicable.

*These data were not collected within the present study but are from a separate collaborative research project between MRCP(UK) and MRCGP.

Table 4| Mean (SD) marks of UK and PLAB graduates at their first attempt at the various parts of MRCP(UK) and MRCGP exams. (All differences are significant with P<0.001)

Exam (first attempt)	UK graduates		PLAB graduates		Effect size
	Mean (SD) marks	No of graduates	Mean (SD) marks	No of graduates	
MRCP(UK)					
Part 1	0.73 (10.12)	18 352	-8.73 (10.75)	7823	$\Delta = -0.94$
Part 2	6.41 (7.49)	14 094	-0.41 (6.91)	5133	$\Delta = -0.91$
PACES*	1.15 (5.34)	14 376	-6.34 (6.32)	4386	$\Delta = -1.40$
MRCGP					
AKT	19.02 (15.04)	12 152	3.78 (16.23)	3160	$\Delta = -1.01$
CSA	13.44 (9.93)	5977	-4.61 (10.66)	1388	$\Delta = -1.82$

*PACES marks were converted to the "old PACES" scoring system

Table 5| Summary of estimated change in the pass mark of PLAB Part 1 and PLAB Part 2 to produce equivalence, using the two separate methods described in the text

	PLAB Part 1			PLAB Part 2	
	MRCP(UK) Part 1	MRCP(UK) Part 2	MRCGP AKT	MRCP(UK) PACES	MRCGP CSA
Method 1: Equivalence of medians	+25	+32	+35	> +18	+10
Method 2: Comparison with UK medical schools (midpoint value)	+20 to +24 (+22)	+30 to +34 (+32)	+15 to +19 (+17)	+16 to +18 (+17)	+16 to +18 (+17)

Values are indicated as marks relative to the current pass mark and are in raw marks on the scales of the examinations themselves. The marking scales for PLAB Part 1 and PLAB Part 2 are not comparable.

Figures

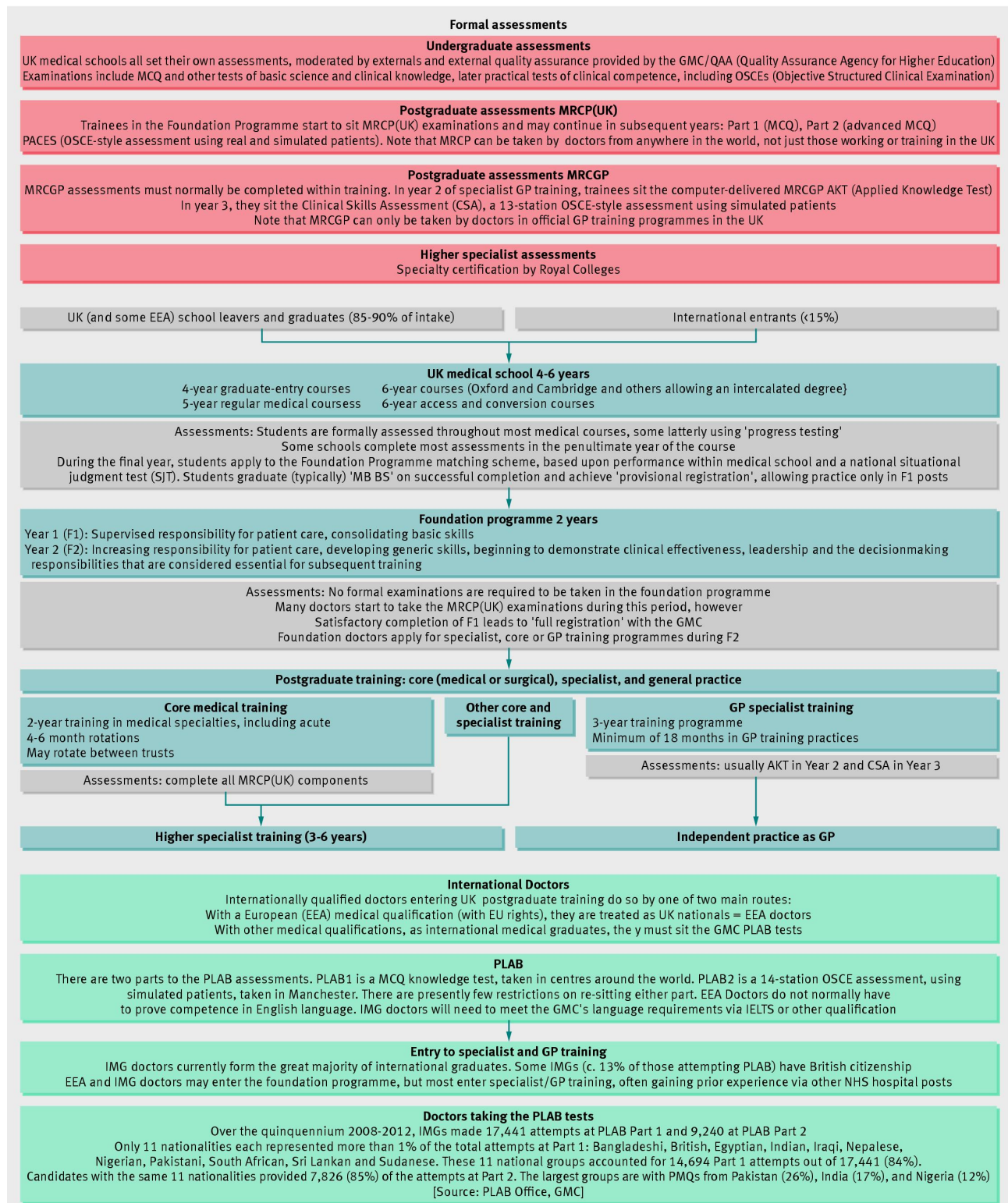


Fig 1 Summary of the selection, assessment and training of UK doctors (undergraduate and postgraduate) and international medical graduates (postgraduate only)

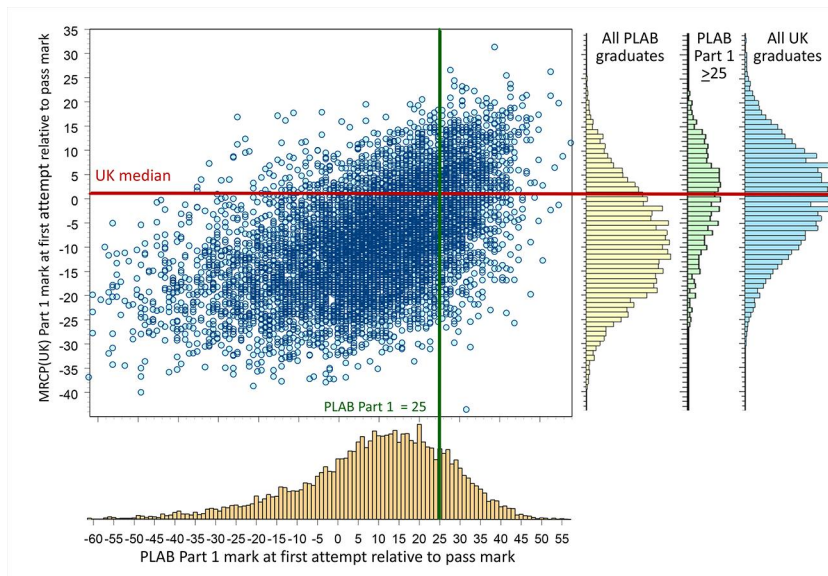


Fig 2 Example of derivation of an equivalent median score for PLAB and UK graduates for MRCP(UK) Part 1. (For explanation, see text)

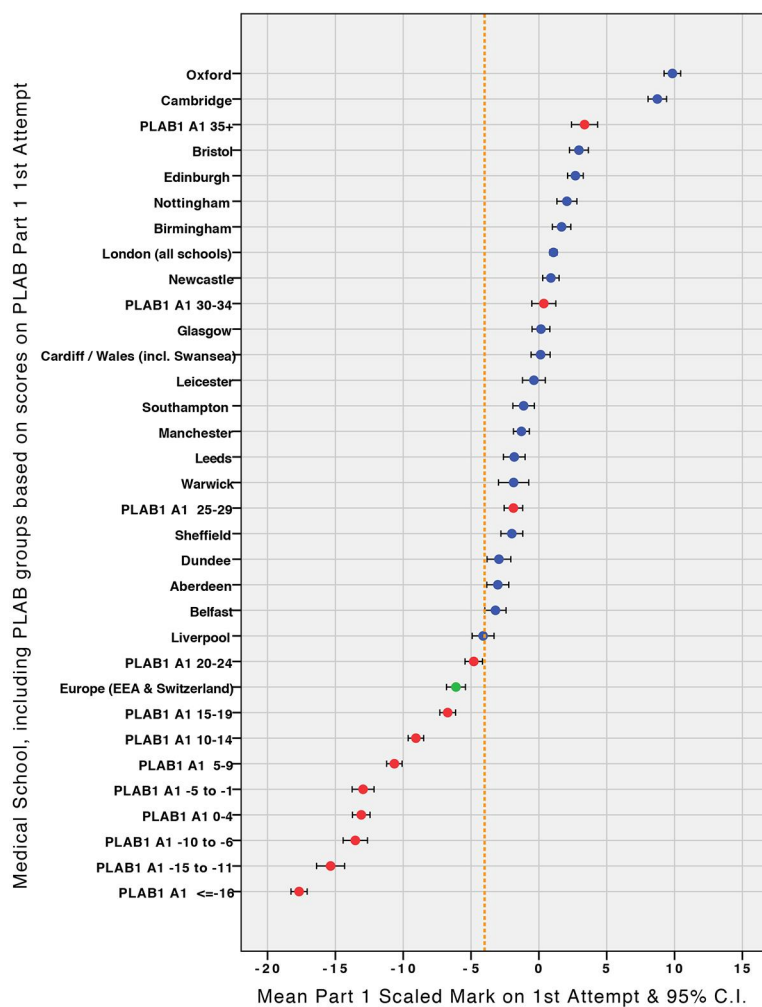


Fig 3 Mean performance on the MRCP(UK) Part 1 of graduates of UK universities (blue points) in relation to the performance of PLAB graduates divided into 12 groups according to PLAB Part 1 mark at first attempt (red points). EEA graduates are shown as a green point. MRCP(UK) data does not identify individual schools within the University of London

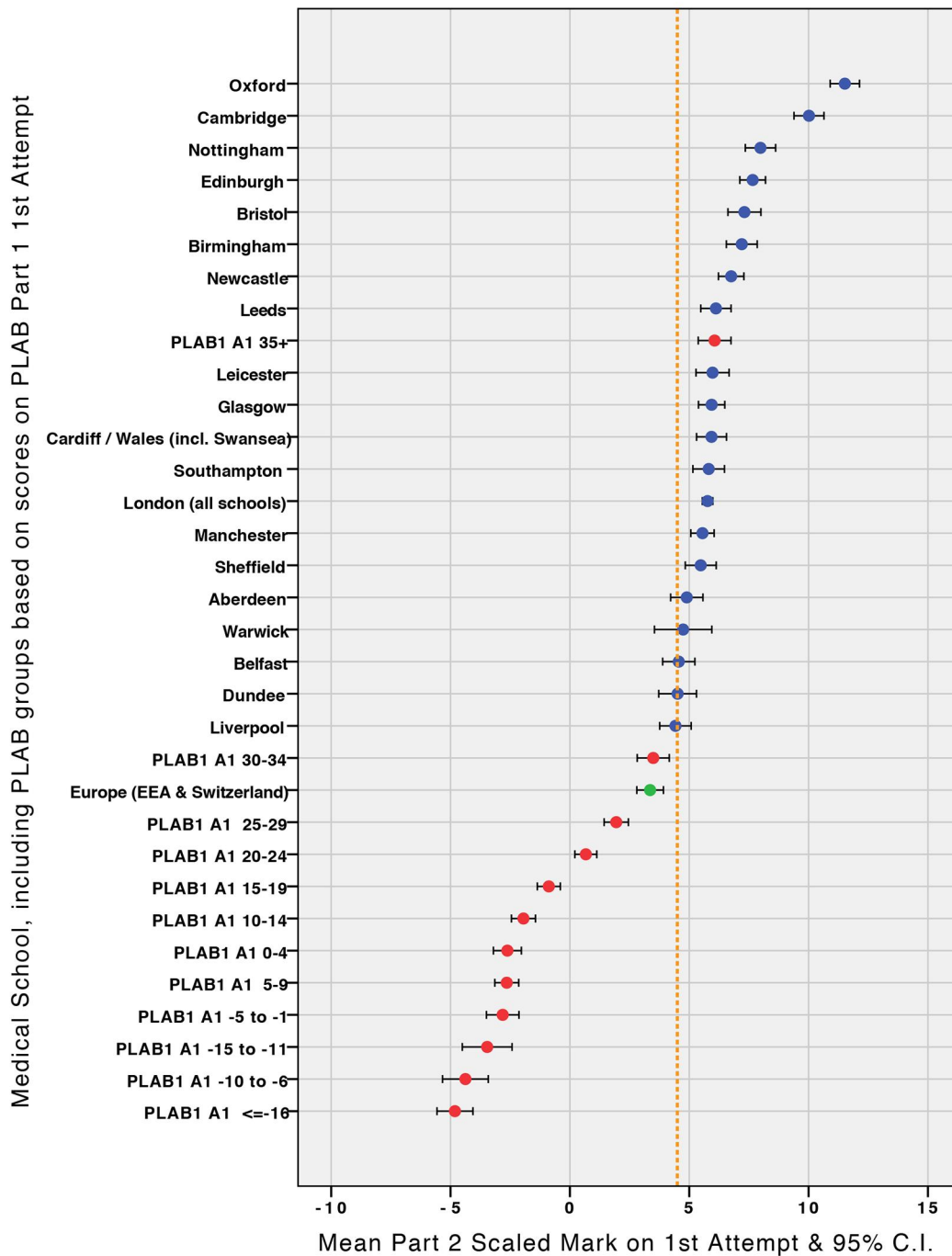


Fig 4 Mean performance on the MRCP(UK) Part 2 of graduates of UK universities (blue points) in relation to the performance of PLAB graduates divided into 12 groups according to PLAB Part 1 mark at first attempt (red points). EEA graduates are shown as a green point. MRCP(UK) data does not identify individual schools within the University of London.

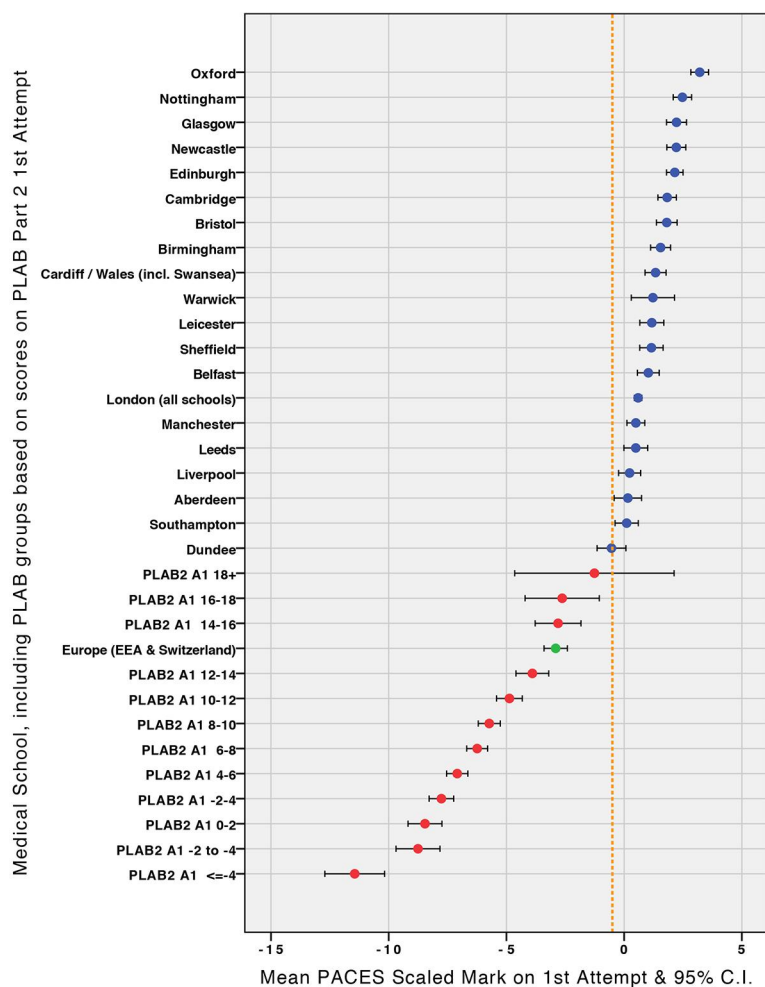


Fig 5 Mean performance on the MRCP(UK) PACES of graduates of UK universities (blue points) in relation to the performance of PLAB graduates divided into 12 groups according to PLAB Part 2 mark at first attempt (red points). EEA graduates are shown as a green point. MRCP(UK) data does not identify individual schools within the University of London.

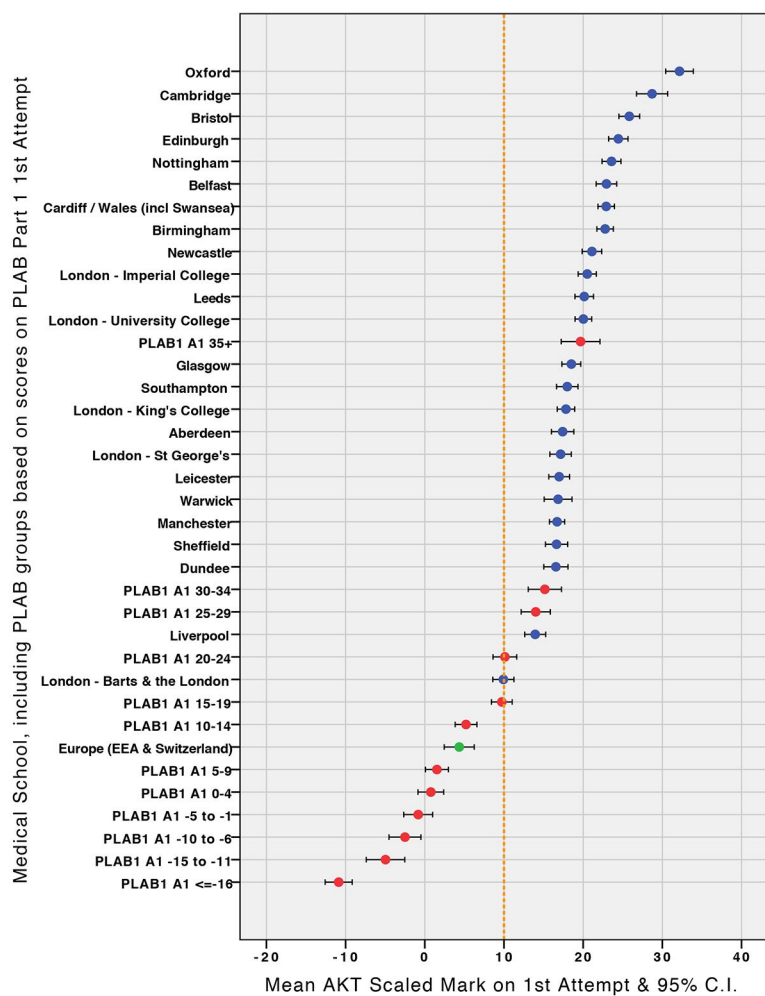


Fig 6 Mean performance on the MRCGP AKT of graduates of UK medical schools (blue points) in relation to the performance of PLAB graduates divided into 12 groups according to PLAB Part 1 mark at first attempt (red points). EEA graduates are shown as a green point.

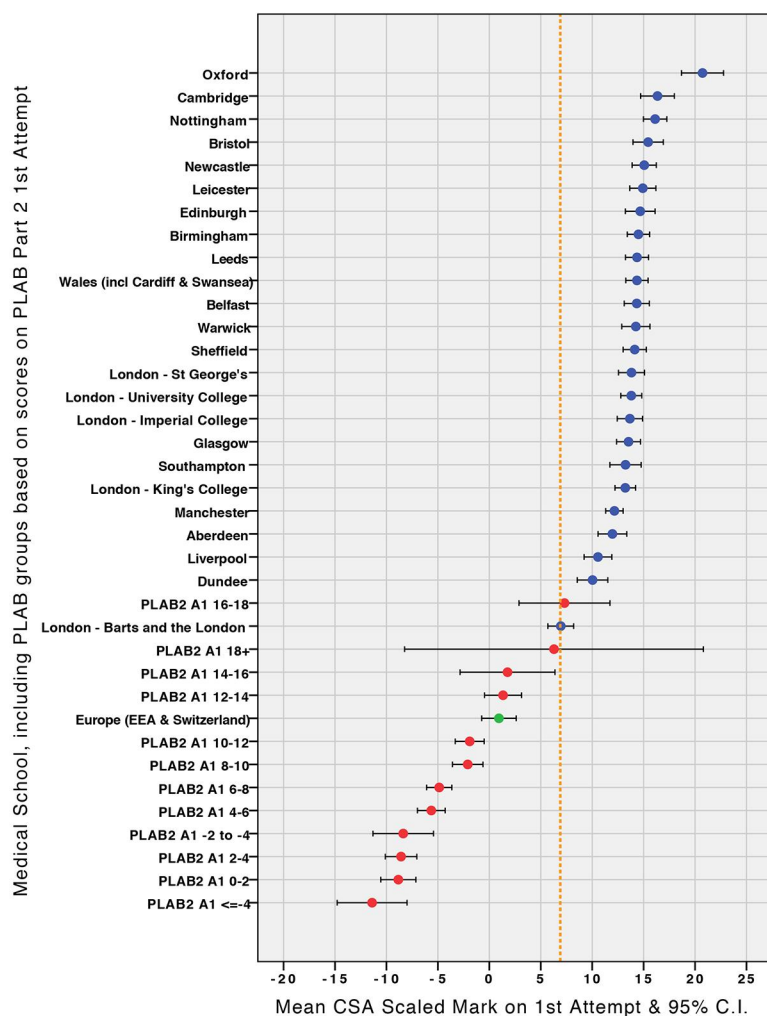


Fig 7 Mean performance on the MRCGP CSA of graduates of UK medical schools (blue points) in relation to the performance of PLAB graduates divided into 12 groups according to PLAB Part 2 mark at first attempt (red points). EEA graduates are shown as a green point.

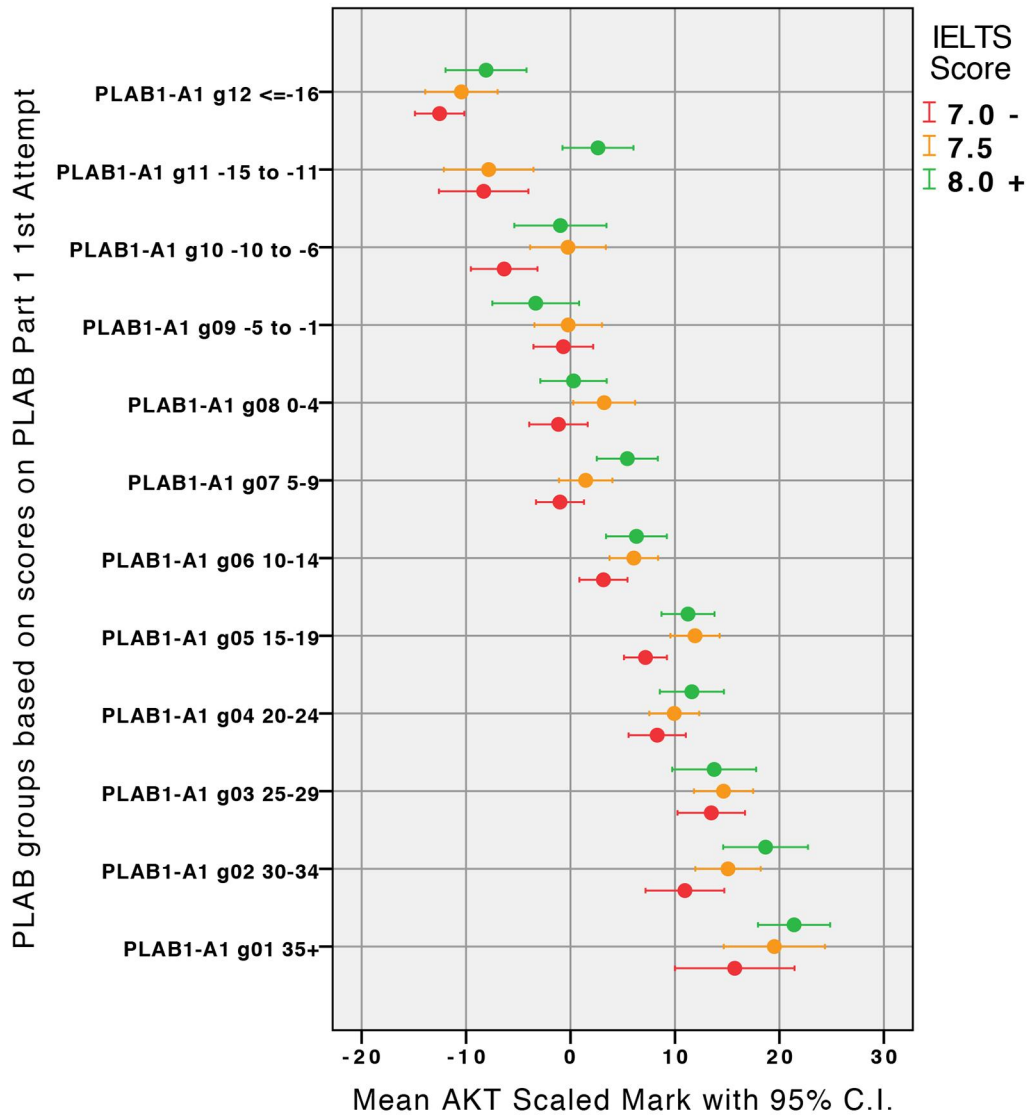


Fig 8 IELTS and MRCGP AKT performance of PLAB graduates. Performance at the AKT (horizontal axis) in relation to performance at PLAB Part 1 (first attempt) by IELTS score (red ≤7.0, orange 7.5, green ≥8.0)

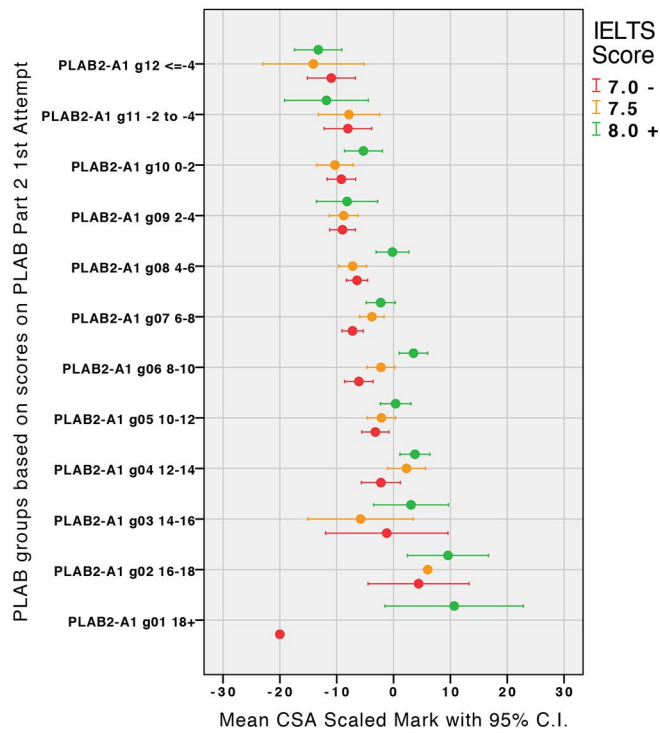


Fig 9 IELTS and MRCGP CSA performance of PLAB graduates. Performance at the CSA (horizontal axis) in relation to performance at PLAB Part 2 (first attempt) by IELTS score (red ≤7.0, orange 7.5, green ≥8.0)