

PLAB Annual report

Part 1 and Part 2

2015 – 2016

Overview

This report outlines the performance of candidates, assessments, assessment items and assessors across the PLAB Part 1 and 2 examinations that were administered between September 2015 and June/July 2016. During this period, 34 Part 1 tests and 55 Part 2 OSCEs were administered. Part 1 was taken by 2667 candidates with 87.5% taking Part 1 once, 10.8% twice, 1.5% three times and 0.3% four times during this period. Outside of 2015-2016, 7% of candidates had three or more attempts at passing Part 1; 5.5% had two attempts; and 13.2% had one additional attempt. Of the total Part 1 candidature, 65.5% were taking Part 1 for the first and only time in 2015-2016.

Part 2 was taken by 2139 candidates with 76.5% completing Part 2 once during 2015-2016, 20% twice, 3.1% three times, and 0.4% four times. Over the course of 2015-2016, a total of 38,000 stations were delivered, and assessed by 162 examiners. The contribution of each examiner varied greatly, with some examiners assessing several hundred candidates (one examiner assessed nearly 1000 candidates), and others assessing for one Part 2 administration only (approximately 40 candidates). Of the 4806 candidate entries processed across PLAB between September and June/July, 26.6% (n=1278) were for candidates taking both parts. Of these, 66.4% (n=849) successfully completed both components in the one year.

The report is organised in five sections:

Section 1: Overview of Part 1 performance **Page 3**

Section 1 reviews the pass rates, mean scores and reliability coefficients for Part 1; an analysis of test performance by selected candidate demographic characteristics; and the relationship between Part 1 and candidates' component-level IELTS scores.

Section 2: Part 1 item analysis **Page 16**

Section 2 records the number of uses of each item and the facility and discrimination values; an analysis of item difficulty by selected blueprint classifications (topic, skill and domain); and an analysis of the relationship between perceived difficulty (Angoff scores) and observed difficulty (actual scores).

Section 3: Overview of Part 2 performance **Page 19**

Section 3 includes analysis of Part 2 scores, pass rates, and number of stations passed for the whole cohort, by selected demographic characteristics, and by IELTS scores. The reliability coefficient for each OSCE administration is also provided.

Section 4: Part 2 station analysis **Page 33**

Section 4 records the mean score and percentage of passes and fails by each station, including analysis of mean score and pass rate by the station skill classification.

Section 5: Examiner performance **Page 34**

Section 5 records the marking trends for each Part 2 examiner in 2015-2016, including scores awarded, the proportion of station passes and fails, and the distribution of global judgements.

Key findings

- Across all candidates examined in 2015-2016, 72.3% (n=2207) achieved a pass for Part 1, and 70.5% (n=1827) achieved a pass for Part 2.
- The pass rate among first take candidates was 78.8% (n=1731) for Part 1, and 71.1% (n=1318) for Part 2.
- There was some variability in the spread of pass rates across each individual test administration for Parts 1 and 2, but half of all pass rates were between 63% and 80% for both components.
- The proportion of IMG and UK candidates passing Parts 1 and 2 on first attempt was comparable at 79%.
- On the whole, candidates aged 32 and under tended to produce the best performances for Part 1 and Part 2, and achieved the highest pass rate. The highest mean score for Part 1 was achieved by a small group of candidates aged 52 and over.
- Candidates' gender explained a small proportion of the observed score variance, but had no real effect on the distribution of scores for Part 1 or 2.
- Candidates' ethnic origin was not influential in explaining any of the observed score variance for Part 1 or 2 scores.
- Of the 160 examiners for Part 2, 68% (n=108) awarded scores within one standard deviation of the mean; 16% (n=26) awarded scores that were higher than this, and 16% (n=26) awarded scores that were lower than this.
- Candidate entries were received from 95 nationalities. Of the nationalities with more than ten candidates, the most successful for Part 1 were the candidates from Myanmar and Australia. For Part 2, the Australian candidates outperformed candidates from all other nationalities with a pass rate of 94%.
- The Part 1 items most likely to be answered correctly were mostly within the *urology* and *developmental problems* topic areas. The items most likely to be answered incorrectly were mostly within the *musculoskeletal* and *renal* topic areas.
- Across the Part 2 stations, the practical skills stations produced the highest scores, but also the lowest pass rate. The most difficult station from all those delivered in 2015-2016 was a station on epilepsy. The station producing the highest score focused on knee examination.

Notes on data

- The data for Part 1 cover September 2015 to end of June 2016. An additional 382 candidates were examined for Part 1 in July/August 2016 but are not included in this report. The data for Part 2 cover the OSCEs delivered between September 2015 and mid-July 2016.
- The total number of candidates in each table may not be consistent due to missing data in the variables included in the analysis.
- Data are based on all candidates except where it is stated that the focus is on first take candidates only.
- The attempt number is not recorded for Part 2. As such, the number of attempts for Part 2 has been calculated based on number of entries between September 2015 and July 2016. Therefore, the number of attempts for Part 2 may be underestimated as candidates may also have attempted Part 2 in 2014-2015 or earlier.

Section 1: Part 1 performance

Section 1 explores total test scores and pass rates for the Part 1 examination by a number of test and candidate characteristics.

Section 1a: Part 1 results

Table 1 compares the pass and fail rates for each Part 1 test administration. The rates are variable across the different administrations, from the lowest pass rate of 47% to the highest of 90%. Across all administrations, 50% of all pass rates fell between 63% and 80%. The observed variability is likely to be a product of cohort size combined with minor fluctuations in candidate ability, and the distribution of item difficulty, across test administrations. The difficulty and discrimination values for each item are recorded in Appendix 1.

Table 1 also shows the Cronbach's alpha coefficients. For completeness, a coefficient is shown for each Part 1 administration. However, the number of candidates for some administrations is small and alpha should therefore be treated with some caution. Across all administrations, 50% of the alpha coefficients were between .91 and .94.

The average coefficient for 2015-2016 was 0.92, which exceeds the standard traditionally conceived as being required for high stakes examinations (0.80). One administration produced an alpha that was within an acceptable range, but below the 0.80 threshold (highlighted below). This may be partly explained by the small number of candidates taking this particular test (n=10), but not exclusively so, as another test with ten candidates produced an alpha of 0.97.

Table1. Part 1 pass/fail rate and Cronbach's alpha by individual test administration.

Part 1	Fail (%)	Pass (%)	Total (count)	Alpha
203439	40.0	60.0	10	.974
203440	37.5	62.5	40	.926
203441	47.4	52.6	19	.947
203442	24.9	75.1	197	.937
203443	25.0	75.0	36	.941
203444	29.2	70.8	24	.933
203445	25.0	75.0	20	.864
203446	30.4	69.6	56	.916
203447	14.3	85.7	21	.880
203448	52.9	47.1	17	.943
203449	26.3	73.7	38	.908
203450	23.5	76.5	17	.961
203451	18.8	81.3	32	.943
203452	38.1	61.9	42	.946
203453	34.3	65.7	99	.928
203454	38.3	61.7	384	.937
203604	42.9	57.1	14	.962
203605	37.0	63.0	54	.939

Part 1	Fail (%)	Pass (%)	Total (count)	Alpha
203606	26.3	73.7	19	.926
203607	22.9	77.1	96	.910
203608	17.5	82.5	194	.928
203609	23.7	76.3	38	.913
203610	10.5	89.5	38	.804
203611	19.0	81.0	21	.897
203612	15.9	84.1	88	.889
203613	20.0	80.0	10	.699
203614	21.3	78.7	47	.910
203615	11.1	88.9	18	.897
203616	20.4	79.6	421	.919
203617	15.4	84.6	52	.921
203618	37.5	62.5	32	.913
203619	38.1	61.9	21	.933
203620	33.5	66.5	457	.932
203663	27.2	72.8	382	.932
Average	27.7	72.3	100.0	.915
Total	847	2207	3054	

Table 2 shows the proportion of international medical graduates (IMG) and UK candidates passing and failing the Part 1 examination in 2015-2016 by the number of the attempt. Across all 3,054 candidates taking the Part 1 examination

in this period, 72% were taking the test for the first time; 14.5% for the second time; 6% for the third time; and the remaining 7.5% were taking the test for the fourth or more time. Of those taking the test for the first time, 79% passed, gaining an average score of 73%. Of those taking the test for the second time, 61.5% passed, gaining an average score of 70%. As we may expect, the pass rate and average score among the passing candidates decreased with each subsequent attempt, before the pass rate stabilises at around 41/42% for candidates taking Part 1 five or more times. A total of 56 candidates were recorded as having seven or more attempts at Part 1; of these, four out of ten went on to pass Part 1.

While the cohort sizes are notably different for IMG (n=2015) and UK (n=159) candidates, the proportion of each cohort passing on first attempt was consistent during this period at 78/79%. Indeed, it is not until the fifth attempt – where the numbers of candidates are small – that the pass rate for each cohort diverges. During this period, IMG candidates were more likely to pass on first take than the UK candidates (79% and 72% respectively). Across all attempts the IMG candidates achieved a pass rate of 72.6%, compared with 69.8% among the UK candidates.

The final two columns show that there is some consistency in the mean score of failing candidates irrespective of the number of the attempt (at around 52/53%). In contrast, the mean score of passing candidates decreases with each attempt, from a score of 73.3% on first attempt to a score of 66.5% on the seventh or more attempt.

Table 2. Part 1 pass/fail rate by selected nationality world regions and number of attempts¹.

Attempt		IMG			UK			Total			Average Part 1 score (%)	
		Fail	Pass	Total N / %	Fail	Pass	Total N / %	Fail	Pass	Total N / %	Fail	Pass
1 st	Result (%)	21.2	78.8	2015	22.0	78.0	159	21.2	78.8	2198	52.8	73.3
	In region (%)	56.6	79.0	72.9	46.7	71.7	64.1	55.1	78.4	72.0		
2 nd	Result (%)	38.0	62.0	395	39.0	61.0	41	38.5	61.5	442	52.2	70.2
	In region (%)	19.8	12.2	14.3	21.3	14.5	16.5	20.1	12.3	14.5		
3 rd	Result (%)	44.1	55.9	161	44.4	55.6	18	44.8	55.2	183	53.0	69.8
	In region (%)	9.4	4.5	5.8	10.7	5.8	7.3	9.7	4.6	6.0		
4 th	Result (%)	51.9	48.1	79	50.0	50.0	10	51.6	48.4	91	53.8	69.5
	In region (%)	5.4	1.9	2.9	6.7	2.9	4.0	5.5	2.0	3.0		
5 th	Result (%)	60.0	40.0	45	42.9	57.1	7	57.1	42.9	56	52.9	68.5
	In region (%)	3.6	.9	1.6	4.0	2.3	2.8	3.8	1.1	1.8		
6 th	Result (%)	56.5	43.5	23	50.0	50.0	4	57.1	42.9	28	53.2	67.7
	In region (%)	1.7	.5	.8	2.7	1.2	1.6	1.9	.5	.9		
≥7 th	Result (%)	56.5	43.4	46	66.7	33.3	9	58.9	41.1	56	55.1	66.5
	In region (%)	3.4	1.0	1.7	8.0	1.7	3.6	3.9	1.0	1.8		
Total	Count	756	2008	2764	75	173	248	847	2207	3054	52.9	72.5
	Result (%)	27.4	72.6	100.0	30.2	69.8	100.0	27.7	72.3	100.0		

Table 3 shows the descriptive statistics for each administration of the Part 1 test. For the first five data columns, red shaded cells indicate a measure that is more than one standard deviation above the mean for that variable (tendency

¹ Data not shown for EEA (n=22) or Unspecified (n=20) candidates due to small numbers.

towards higher scores); while blue shading indicates a measure that is more than one standard deviation below the mean for that variable (tendency towards lower scores). The tests with multiple instances of blue or red shading represent the extremes of Part 1 performance. For example, test 203448 produced a mean score of 118.7 compared with test 203451 which produced a mean score of 144.5; a difference of nearly 26 items. Table 2, above, shows that the pass rate was also markedly different for these two tests, with 47.1% passing test 203448 and 81.3% passing test 203451. While it is important to note the difference, there may be a number of explanations including:

1. Difference in size of candidature: test 203448 comprised only 17 candidates, compared with 32 for 203451;
2. Differences in the spread of ability between the two cohorts;
3. The proportion of first take and retake candidates within the cohort; 23.5% and 6.5% of each test cohort, respectively, were first take candidates. Given the trends shown in Table 1, it appears likely that the higher proportion of re-take candidates may have had an effect of reducing the mean score for test 203448.

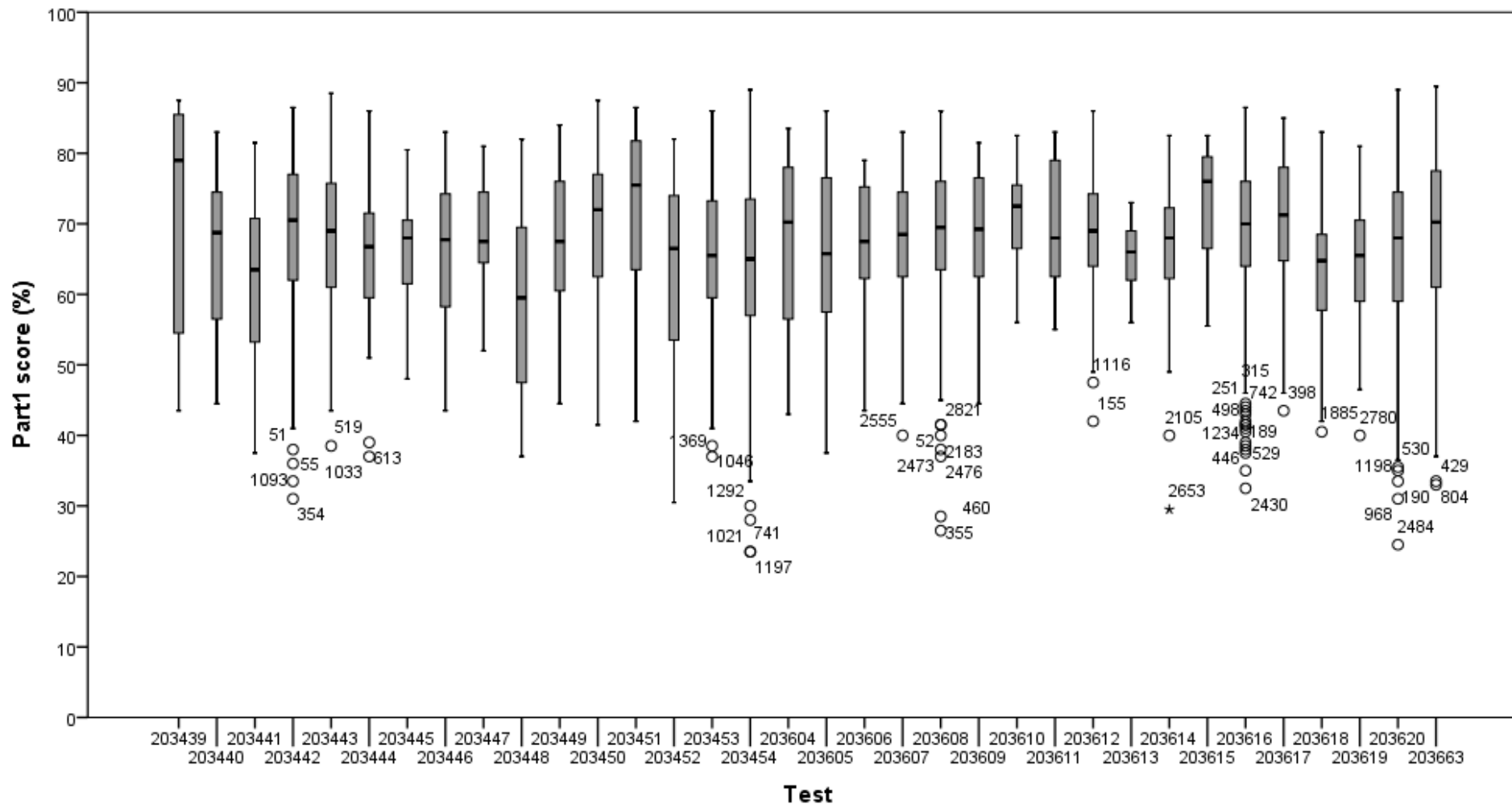
The Cronbach's alpha coefficient for both tests was identical, suggesting that the outcomes for one test were no more or less reliable than the outcomes for the other test.

Table 3. Descriptive scores for Part 1 by test administration (%).

Part 1 test	Mean score	95% Confidence Interval for Mean		5% Trimmed Mean	Median	Std. Deviation	Min	Max	Range	IQR
		Lower	Upper							
203439	70.3	58.2	82.4	70.8	79.0	16.9	43.5	87.5	44.0	32.5
203440	66.2	62.7	69.6	66.4	68.8	10.8	44.5	83.0	38.5	18.5
203441	61.5	55.2	67.9	61.8	63.5	13.2	37.5	81.5	44.0	22.0
203442	68.5	66.9	70.1	69.1	70.5	11.4	31.0	86.5	55.5	16.0
203443	66.8	62.7	70.8	67.2	69.0	12.0	38.5	88.5	50.0	15.5
203444	64.6	59.7	69.4	65.0	66.8	11.5	37.0	86.0	49.0	12.5
203445	66.8	63.0	70.6	67.0	68.0	8.1	48.0	80.5	32.5	10.5
203446	66.0	63.2	68.8	66.3	67.8	10.3	43.5	83.0	39.5	16.5
203447	68.0	64.2	71.7	68.1	67.5	8.3	52.0	81.0	29.0	11.5
203448	59.3	52.6	66.0	59.3	59.5	13.0	37.0	82.0	45.0	24.0
203449	67.4	64.2	70.6	67.7	67.5	9.8	44.5	84.0	39.5	15.5
203450	69.0	61.7	76.4	69.5	72.0	14.3	41.5	87.5	46.0	20.0
203451	72.2	68.1	76.4	73.0	75.5	11.5	42.0	86.5	44.5	18.5
203452	63.9	59.8	67.9	64.6	66.5	12.9	30.5	82.0	51.5	21.0
203453	65.1	62.9	67.3	65.4	65.5	10.9	37.0	86.0	49.0	14.5
203454	64.2	63.0	65.4	64.7	65.0	12.0	23.5	89.0	65.5	16.5
203604	65.7	57.1	74.3	66.0	70.3	14.9	43.0	83.5	40.5	25.0
203605	65.3	62.1	68.6	65.7	65.8	11.8	37.5	86.0	48.5	19.5
203606	65.9	60.7	71.1	66.4	67.5	10.8	43.5	79.0	35.5	15.5
203607	67.3	65.4	69.2	67.7	68.5	9.5	40.0	83.0	43.0	12.5
203608	68.5	67.1	70.0	69.3	69.5	10.3	26.5	86.0	59.5	12.5
203609	68.2	65.0	71.4	68.7	69.3	9.7	44.5	81.5	37.0	14.5
203610	70.9	68.8	73.1	71.1	72.5	6.4	56.0	82.5	26.5	9.0
203611	69.8	65.8	73.8	69.9	68.0	8.8	55.0	83.0	28.0	17.0
203612	68.8	67.0	70.6	69.1	69.0	8.6	42.0	86.0	44.0	10.5
203613	65.7	61.8	69.5	65.8	66.0	5.4	56.0	73.0	17.0	8.5
203614	66.6	63.7	69.4	67.4	68.0	9.7	29.5	82.5	53.0	10.5
203615	72.7	68.6	76.7	73.1	76.0	8.2	55.5	82.5	27.0	13.5
203616	68.6	67.6	69.5	69.2	70.0	9.9	32.5	86.5	54.0	12.0
203617	70.1	67.4	72.9	70.7	71.3	9.9	43.5	85.0	41.5	13.5
203618	62.9	59.3	66.6	63.1	64.8	10.2	40.5	83.0	42.5	11.5
203619	64.2	59.0	69.3	64.6	65.5	11.3	40.0	81.0	41.0	16.5
203620	66.1	65.0	67.1	66.6	68.0	11.0	24.5	89.0	64.5	16.0
203663	68.7	67.5	69.8	69.1	70.3	11.1	33.0	89.5	56.5	16.5
Average	66.9	63.1	70.7	67.3	68.6	10.7	40.4	84.0	43.6	15.8
SD	2.88	3.93	3.49	2.96	3.66	2.29	8.58	3.36	10.89	5.01

Chart 1, below, shows the spread of scores awarded for each test. As in shown in Table 2, there is considerable variability in the number of candidates per administration. As such, the observed variability in the spread of scores does not, in itself, indicate that any one administration was more difficult or easy than another, as there may be other explanatory factors.

Chart 1. Distribution of Part 1 scores (%) by test administration.



Section 1b: Part 1 performance by candidates' demographic characteristics

The next section explores Part 1 performance by the following selected demographic characteristics of trainees within each test cohort: gender, age, ethnic origin, country of primary medical qualification, and nationality. This section focuses on first take candidates only, and excludes from the analysis all re-take candidates. The total number of candidates per administration and by each characteristic is therefore determined by the proportion of first take candidates combined with the availability or completeness of demographic profiles for individual candidates. The comparisons of average scores and pass rates by the various demographic characteristics should be treated with caution as each demographic cohort (e.g. candidates from Australia, or those aged between 21 and 26) will have taken different, but parallel, versions of Part 1.

Table 4 shows the Part 1 pass and fail rate and descriptive scores by gender for all first take candidates in 2015-2016. The table shows that the differences between the gender cohorts are marginal. Of the 467 candidates who failed their first take of Part 1, there were slightly more fails among female candidates (+1.4%). Within each of the gender groups, female candidates gained marginally more passes than the male candidates (+1.6%). On average, female candidates also gained a marginally higher mean score, but this equates to an average of just over one additional item correct.

Table 4. Pass/fail rate and descriptive scores by gender (first take).

Part 1 Result		Gender		Total/ average
		Male	Female	
Fail	% result	49.3	50.7	467
	% gender	22.1	20.5	21.2
Pass	% result	47.0	53.0	1731
	% gender	77.9	79.5	78.8
Total	% result	47.5	52.5	2198
	% gender	100.0	100.0	100.0
Mean		68.6	69.3	69.0
SD ²		10.7	10.4	10.5
Min		23.5	31.0	23.5
Max		87.5	89.5	89.5
Range		64.0	58.5	66.0
IQR ²		13.5	13.0	13.1

Table 5 shows the pass/fail rate and descriptive scores by age band³. The youngest candidates, those aged between 21 and 26, were, on average, most successful at Part 1, achieving a pass rate of 84% and a mean score of 70.8%. A small group of candidates aged 52 or above were the next most successful group, with a pass rate of 78.6% and a higher mean score of 73%, followed very closely by the 27-32 age group (78.5% pass rate and a mean score of 68.9%). Interestingly, the candidates aged 52 and over produced scores in a much smaller range than the other groups, ranging between 105 and 169 items correct (range=64), compared with the youngest age group who scored between 47 and 177 items correct (range=130).

² Standard deviation; Interquartile range.

³ Six age bands were created; each covering six years, with the exception of the most mature group who were aged 52-65. This group comprised far smaller numbers than the other groups (n=34; 1.1% of all Part 1 candidates).

Table 5. Pass/fail rate and descriptive scores by age (first take).

Part 1 result		Age						Total/ average
		21-26	27-32	33-38	39-44	45-51	52+	
Fail	% result	18.8	51.6	21.8	5.4	1.7	0.6	467
	% age	15.9	21.5	27.3	25.3	23.5	21.4	21.3
Pass	% result	26.9	50.9	15.7	4.3	1.5	0.6	1730
	% age	84.1	78.5	72.7	74.7	76.5	78.6	78.7
Total	Count	554	1122	374	99	34	14	2197
	% result	25.2	51.1	17.0	4.5	1.5	0.6	100.0
Mean		70.8	68.9	67.0	68.0	65.8	73.1	69.0
SD		10.2	10.7	10.1	11.2	9.0	10.4	10.5
Min		23.5	30.5	29.5	31.0	41.5	52.5	23.5
Max		88.5	89.5	87.5	85.0	80.0	84.5	89.5
Range		65.0	59.0	58.0	54.0	38.5	32.0	66.0
IQR		12.5	13.5	13.5	14.5	10.5	17.0	13.1

Table 6 shows the pass/fail rate and mean scores by candidates' ethnic origin. The largest ethnic group taking Part 1 in 2015-2016 comprised Asian or Asian British candidates (53.8% of total candidature). The average fail rate across all first take candidates was 21.2%; the fail rate for Asian or Asian British (n=1182) and Black or Black British (n=458) candidates was marginally above this at 22.4% and 22.3% respectively. The lowest fail rate was observed for the much smaller group of candidates who classified themselves as being of Mixed ethnic origin (15.6%, n=32).

Table 6. Pass/fail rate and descriptive scores by ethnicity (first take).

Ethnic origin	Part 1 result						Part 1 score (%)					
	Fail		Pass		Total		Mean	SD	Min	Max	Range	IQR
	% ethnic origin	% result	% ethnic origin	% result	N	%						
Asian or Asian British	22.4	56.7	77.6	53.0	1182	53.8	68.9	10.9	23.5	89.0	65.5	14.0
Black or Black British	22.3	21.8	77.7	20.6	458	20.8	68.1	9.9	33.0	86.5	53.5	11.5
Mixed	15.6	1.1	84.4	1.6	32	1.5	69.7	11.4	39.0	86.5	47.5	14.6
Not stated	15.7	3.4	84.3	5.0	102	4.6	69.5	10.7	32.5	89.5	57.0	14.1
Other Ethnic Groups	19.9	10.7	80.1	11.6	251	11.4	69.7	10.2	24.5	87.0	62.5	14.0
Unspecified	17.9	1.1	82.1	1.3	28	1.3	71.6	11.2	38.5	86.0	47.5	16.3
White	16.6	5.1	83.4	7.0	145	6.6	69.9	10.1	40.0	87.0	47.0	14.0
Total/average	21.2	100.0	78.8	100.0	2198	100.0	69.0	69.0	69.0	69.0	69.0	69.0

Table 7 shows the pass/fail rate and mean score for all first take Part 1 candidates by the country in which they achieved their primary medical qualification (PMQ). For completeness, all candidates are shown in the table, including where there are very few entries. The largest cohorts qualified in Pakistan (n=566), Nigeria (n=328), India (n=288), Egypt (n=115) and the Sudan (n=75). Within each of these large cohorts, approximately three-quarters of candidates achieved a pass. The highest pass rate was achieved by candidates qualifying in Myanmar⁴ (97.4%, n=38), closely

⁴ Excluding countries with fewer than 20 candidates.

followed by candidates qualifying in Australia (97%, n=33). The lowest pass rates were observed among candidates qualifying in the Ukraine (57.5%, n=40) and China (62.3%, n=61).

The candidates with British nationality (n=238), achieved their primary medical qualification in a wide range of countries, with the largest numbers achieving qualification in Pakistan (19.7%, n=47), the Sudan (14.7%, n=35) and India (8.8%, n=21).

Table 7. Pass rate and mean scores by country of primary medical qualification (first take).

PMQ	Part 1 result		N	Part 1 score	
	Fail	Pass		Mean	SD
Afghanistan	█	█	█	█	█
Algeria	█	█	█	█	█
Antigua And Barbuda	█	█	█	█	█
Argentina	█	█	█	█	█
Armenia	40.0	60.0	5	63.5	18.6
Australia	3.0	97.0	33	74.4	6.8
Azerbaijan	█	█	█	█	█
Bahrain	19.4	80.6	31	66.9	6.6
Bangladesh	13.8	86.2	65	71.2	10.6
Barbados	█	█	█	█	█
Belarus	0.0	100.0	4	71.1	7.6
Bosnia And Herzegovina	█	█	█		
Brazil	25.0	75.0	8	71.4	9.7
Bulgaria	█	█	█	█	█
Cameroon	█	█	█	█	█
Canada	█	█	█		
Cayman Islands	█	█	█	█	█
Chile	█	█	█		
China	37.7	62.3	61	64.2	12.6
Colombia	█	█	█	█	█
Cuba	16.7	83.3	6	67.3	8.7
Curacao	█	█	█	█	█
Czech Republic	33.3	66.7	6	70.7	8.6
Dominican Republic	█	█	█		
Egypt	20.0	80.0	115	70.2	11.0
El Salvador	█	█	█		
Fiji	█	█	█	█	█
Gambia	█	█	█		
Georgia	█	█	█	█	█
Ghana	12.5	87.5	16	74.3	9.3
Grenada	13.3	86.7	15	70.1	9.2
Guinea	█	█	█		

PMQ	Part 1 result		N	Part 1 score	
	Fail	Pass		Mean	SD
Kyrgyzstan	█	█	█	█	█
Lebanon	█	█	█	█	█7
Libya	4.8	95.2	21	72.0	7.9
Lithuania	█	█	█	█	█
Malawi	█	█	█		
Malaysia	16.7	83.3	6	70.4	7.3
Mauritius	25.0	75.0	8	64.6	13.6
Mexico	16.7	83.3	6	65.2	4.6
Moldova	█	█	█	█	█
Myanmar	2.6	97.4	38	75.7	6.6
Nepal	16.7	83.3	30	71.1	10.2
New Zealand	0.0	100.0	10	73.1	6.9
Nigeria	22.9	77.1	328	67.8	9.7
Oman	25.0	75.0	4	66.8	13.9
Pakistan	20.0	80.0	566	69.5	10.3
Peru	█	█	█		
Philippines	28.6	71.4	14	67.5	14.7
Poland	37.5	62.5	8	65.9	9.8
Romania	45.5	54.5	11	61.8	15.0
Russia	35.0	65.0	40	65.5	12.1
Saba	█	█	█		
Saint Kitts And Nevis	25.0	75.0	4	69.0	9.3
Saint Lucia	█	█	█		
Saudi Arabia	33.3	66.7	6	60.3	19.9
Serbia	█	█	█		
Singapore	█	█	█		
Slovakia	█	█	█		
South Africa	9.5	90.5	42	70.9	5.9
Sri Lanka	13.6	86.4	22	72.4	7.8
Sudan	14.7	85.3	75	68.9	7.9
Syrian Arab Rep	25.0	75.0	8	69.9	10.1
Taiwan	█	█	█		

PMQ	Part 1 result		N	Part 1 score	
	Fail	Pass		Mean	SD
Guyana	█	█	█	█	█
Hungary	30.0	70.0	10	68.4	13.8
India	26.4	73.6	288	68.2	11.2
Indonesia	33.3	66.7	6	63.7	9.1
Iran	10.0	90.0	10	76.7	8.3
Iraq	16.7	83.3	42	72.0	12.7
Ireland	33.3	66.7	12	63.7	7.4
Israel	█	█	█	█	█
Italy	█	█	█		
Jamaica	30.0	70.0	10	66.9	7.8
Jordan	0.0	100.0	11	75.5	5.3
Kazakhstan	25.0	75.0	4	65.6	8.6
Kenya	9.1	90.9	11	70.8	8.5
Kuwait	█	█	█	█	█

PMQ	Part 1 result		N	Part 1 score	
	Fail	Pass		Mean	SD
Tanzania	25.0	75.0	4	68.3	13.6
Thailand	█	█	█	█	█
Trin & Tobago	█	█	█	█	█
Tunisia	█	█	█		
Turkey	0.0	100.0	8	75.7	5.7
Uganda	█	█	█	█	█
Ukraine	42.5	57.5	40	61.9	11.7
UAE	15.6	84.4	32	69.6	11.2
US	0.0	100.0	4	80.9	7.5
Uzbekistan	█	█	█	█	█
Venezuela	█	█	█	█	█.4
Yemen	█	█	█	█	█
Zambia	█	█	█	█	█
Zimbabwe	8.3	91.7	12	72.0	8.5
Average	21.2	78.8	2198	68.6	9.6

Table 8 shows the pass/fail rate and mean score by candidates' nationality. This table includes 159 British candidates who achieved a pass rate of 78% and an average score of 68.6%, which is just below the average of 69.0%. The 26 Myanmar candidates achieved a pass rate of 100% and the highest average score of 76.2%.

Table 8. Pass/fail rate and descriptive scores by nationality (first take).

Nationality	Part 1 result		N	Part 1 score	
	Fail	Pass		Mean	SD
Afghan	█	█	█		
Albanian	█	█	█		
American	22.7	77.3	22	70.7	10.9
Antiguan	█	█	█	█	█
Argentinian	█	█	█	█	█
Armenian	█	█	█		
Australian	3.3	96.7	30	74.0	7.6
Azerbaijani	█	█	█		
Bahamian	█	█	█	█	█
Bahraini	25.0	75.0	12	67.7	7.9
Bangladeshi	13.6	86.4	44	71.5	10.4
Barbadian	█	█	█		
Belarusian	█	█	█	█	█
Bosnian	█	█	█		
Botswanian	█	█	█		
Brazilian	28.6	71.4	7	71.6	10.4
British	22.0	78.0	159	68.6	11.0
Burmese	9.1	90.9	11	73.5	6.7
Cameroon	40.0	60.0	5	62.2	14.5

Nationality	Part 1 result		N	Part 1 score	
	Fail	Pass		Mean	SD
Lebanese	█	█	█	█	█
Libyan	5.6	94.4	18	72.1	7.3
Malawian	█	█	█		
Malaysia	15.0	85.0	20	69.8	8.3
Maldivian	█	█	█	█	█
Mauritian	16.7	83.3	6	71.2	9.5
Mexican	20.0	80.0	5	65.3	5.2
Moldavia	█	█	█		
Mozambican	█	█	█		
Myanmar	0.0	100.0	26	76.2	6.8
Namibian	█	█	█		
Nepalese	6.3	93.8	32	73.2	9.4
New Zeal	12.5	87.5	8	65.8	8.2
Nigerian	23.0	77.0	357	68.0	9.8
Norwegian	█	█	█		
Omani	█	█	█		
Pakistan	20.9	79.1	555	69.1	10.7
Palestinian	█	█	█	█	█
Peruvian	█	█	█		

Nationality	Part 1 result		N	Part 1 score	
	Fail	Pass		Mean	SD
Canadian	22.0	78.0	50	67.9	7.8
Chilean	█	█	█		
Chinese	50.0	50.0	6	61.2	13.0
Colombia	█	█	█	█	█
Costa Rican	█	█	█		
Cuban	█	█	█		
Egyptian	17.6	82.4	102	70.7	11.0
Emirian	█	█	█		
Fijian	█	█	█		
Filipino	11.1	88.9	9	72.4	11.8
Finnish	█	█	█		
Georgian	█	█	█		
Ghanaian	33.3	66.7	15	65.0	12.2
Grenadian	█	█	█	█	█
Guyanese	█	█	█	█	█
Hungarian	█	█	█		
Indian	28.9	71.1	322	67.3	11.3
Indonesian	50.0	50.0	4	61.9	11.2
Iranian	9.1	90.9	11	74.5	9.8
Iraqi	16.7	83.3	42	73.7	9.6
Irish	█	█	█		
Israeli	█	█	█		
Ivorian	█	█	█		
Jamaican	42.9	57.1	7	67.2	9.3
Japanese	█	█	█		
Jordanian	0.0	100.0	11	74.6	5.7
Kazakhstani	█	█	█		
Kenyan	15.4	84.6	13	69.6	10.5
Kuwaiti	█	█	█	█	█

Nationality	Part 1 result		N	Part 1 score	
	Fail	Pass		Mean	SD
Polish	█	█	█	█	█
Portuguese	█	█	█		
Romanian	█	█	█		
Russian	7.1	92.9	14	73.5	7.8
Salvador	█	█	█		
Saudi Arabian	█	█	█	█	█
Serbian	█	█	█		
Singaporean	0.0	100.0	4	71.0	7.6
South African	19.0	81.0	42	69.3	10.4
Spanish	█	█	█		
Sri Lank	30.0	70.0	40	67.1	11.6
Sudanese	5.3	94.7	57	70.2	6.6
Swedish	█	█	█		
Syrian	22.2	77.8	9	69.4	13.1
Taiwanese	40.0	60.0	5	67.0	12.2
Tanzania	25.0	75.0	4	68.3	13.6
Thai	█	█	█	█	█
Trinidad	0.0	100.0	8	70.2	5.6
Tunisian	█	█	█		
Turkish	0.0	100.0	7	76.8	5.1
Ugandan	50.0	50.0	4	56.4	18.5
Ukrainian	█	█	█	█	█
Unspecified	█	█	█		
Uzbekistani	█	█	█	█	█
Venezuelan	█	█	█	█	█
Yemeni	█	█	█		
Zambian	█	█	█		
Zimbabwean	7.1	92.9	14	71.9	7.8
Average	21.2	78.8	921	68.6	9.6

Univariate analysis of variance (ANOVA) was used to explore the relationship between candidates' Part 1 score and selected demographic characteristics (gender, age and ethnicity). While being mindful that the candidates' scores were obtained on parallel versions of the Part 1 test, this combination of demographic characteristics accounted for 21% of the total observed score variance. The proportion of explained variance was mostly attributable to candidates' age ($p=.00$) and the interaction between age and ethnicity ($p=.04$), although, in both cases, the effect sizes were negligible suggesting that the differences between candidates in terms of these variables have minimal real effect on test outcome ($\text{Eta}^2 = .012$ and $.018$ respectively). There were no statistically significant results by ethnic origin or gender ($p=.99$ and $p=.47$ respectively), or any of the interactions between these variables.

Across the same demographic characteristics, Chi-square was used to explore if the pass and fail results were statistically significantly different from those that would be expected, with all other things between candidates being assumed equal. The findings suggested no significant difference relating to gender: male candidates achieved marginally fewer passes than expected, and females achieved marginally more passes than would be expected, but the findings were not statistically significant. For age, the difference was statistically significant ($p=.002$). The

differences were mostly attributable to the age group 21-26 gaining fewer fails (-25%) and more passes (+7%) than expected; and 33-38 age group with more fails (+28%) and fewer passes (-8%) than would be expected. For ethnic origin, the differences were not statistically significant, with the minor differences between the observed and expected values being attributable to chance alone (data not shown).

Table 9 shows the Part 1 mean score and pass rate across the intersections of age, gender and ethnic origin, as used for the ANOVA and chi-square analyses. Splitting the 2015-2016 cohort by these variables creates small groups and so the data are provided for information only, and not for the purpose of drawing conclusions about micro-level trends in Part 1 performance. The data do show, however, that the trend between age and Part 1 performance (whereby younger candidates tend to outperform more mature candidates, with the exception of the most mature group) is not so clear-cut once we add in variability by gender and ethnic origin.

Table 9. Part 1 score and pass rate for interactions between age, gender and ethnic origin (first take).

Age	Ethnic origin	Pass rate (%)		Score (%)			Total		
		Female	Male	Female	Male	Average	Female	Male	Total
21-26	Asian or Asian British	83.3	84.5	70.5	70.7	70.6	194	148	342
	Black or Black British	83.8	89.5	70.1	70.8	70.4	36	38	74
	Mixed	█	█	█	█	█	█	█	█
	Not stated	93.1	100.0	69.1	75.5	72.4	14	15	29
	Other Ethnic Groups	80.6	86.8	68.5	71.3	70.1	29	38	67
	Unspecified	100.0	100.0	82.8	70.5	76.1	5	6	11
	White	84.0	85.7	67.7	71.5	69.8	11	14	25
	Total	84.1	87.0	70.4	71.2	70.8	293	261	554
27-32	Asian or Asian British	76.3	73.0	69.3	67.6	68.6	359	248	607
	Black or Black British	80.3	83.0	67.8	69.1	68.6	92	147	239
	Mixed	81.8	75.0	70.2	65.1	68.3	14	8	22
	Not stated	82.5	78.6	69.4	67.1	68.3	29	28	57
	Other Ethnic Groups	79.0	76.1	70.6	69.6	70.0	53	71	124
	Unspecified	77.8	100.0	69.9	73.4	71.4	5	4	9
	White	87.5	88.9	71.2	69.6	70.5	37	27	64
	Total	78.5	77.5	69.3	68.4	68.9	589	533	1122
33-38	Asian or Asian British	69.0	55.0	67.9	63.7	66.3	98	60	158
	Black or Black British	70.2	71.4	65.7	65.9	65.8	54	70	124
	Mixed	█		█	█	█	█		█
	Not stated	81.8	71.4	76.5	66.8	70.3	4	7	11
	Other Ethnic Groups	83.3	81.8	69.8	69.6	69.7	20	22	42
	Unspecified	50.0		76.0	54.0	65.0	1	1	2
	White	85.3	92.9	68.7	73.3	70.6	20	14	34
	Total	72.7	68.4	67.7	66.2	67.0	200	174	374
39-44	Asian or Asian British	77.8	70.8	70.9	66.5	68.9	30	24	54
	Black or Black British	80.0	70.0	73.6	67.1	69.3	5	10	15
	Not stated	66.7	50.0	79.0	61.0	67.0	1	2	3
	Other Ethnic Groups	72.7	85.7	54.4	70.1	64.4	4	7	11
	Unspecified			47.5	0.0	47.5	2		2
	White	71.4	75.0	66.6	70.3	68.7	6	8	14

Age	Ethnic origin	Pass rate (%)		Score (%)			Total		
		Female	Male	Female	Male	Average	Female	Male	Total
	Total	74.7	72.5	68.5	67.5	68.0	48	51	99
45-51	Asian or Asian British	92.3	100.0	69.1	67.6	68.5	8	5	13
	Black or Black British	50.0	50.0	0.0	63.6	63.6		6	6
	Mixed	█	█	█	█	█		█	█
	Not stated	50.0		62.0	59.0	60.5	1	1	2
	Other Ethnic Groups	█	█	█	█	█	█	█	█
	Unspecified	100.0		70.5	0.0	70.5	2		2
	White	█	█	█	█	█	█	█	█
	Total	76.5	68.8	67.8	63.5	65.8	18	16	34
52+	Asian or Asian British	█	█	█	█	█	█	█	█
	Other Ethnic Groups	█	█	█	█	█		█	█
	Unspecified	100.0		80.0	0.0	80.0	2		2
	White	█	█	█	█	█	█	█	█
	Total	78.6	75.0	76.5	70.6	73.1	6	8	14
Total		78.7	77.9	69.3	68.6	69.0	1154	1043	2197

Section 1c: Part 1 and IELTS performance

Candidates' IELTS scores can be used as measures of general ability, in addition to language proficiency. As such, it is prudent to check the nature of any relationship between PLAB and IELTS performance, as it is conceivable that poor performance on one could be, at least a partial, predictor of poor performance on the other. That said, the minimum IELTS score is 7.5 (from June 2014)⁵ meaning that there is less variability in the IELTS scores of Part 1 candidates, than among those who took the IELTS test but did not meet the PLAB requirements.

Table 10 shows the correlation between Part 1 score and the scores achieved on each component of the IELTS test. As we may expect, no correlation was found between Part 1 score and the score achieved for the speaking component. Statistically significant correlations were observed for reading, understanding, writing and the overall IELTS score, but in each case the correlation was weak. This suggests that for Part 1, for this cohort of candidates, the IELTS score was not an effective indicator of Part 1 performance.

⁵ Seven candidates from 2015-2016 are carrying IELTS scores that do not meet this standard. Presumably, these candidates registered for PLAB before the new threshold requirement was introduced.

Table 10. Correlations between Part 1 score and IELTS scores (first take, n=2093⁶).

Pearson correlation	IELTS Reading	IELTS Speaking	IELTS Understanding	IELTS Writing	IELTS Overall
PLAB Part 1	.160**	-0.015	.106**	.050**	.096**
IELTS Reading		.279**	.493**	.371**	.739**
IELTS Speaking			.309**	.444**	.667**
IELTS Understanding				.381**	.714**
IELTS Writing					.521**

** . Correlation is significant at the 0.01 level (2-tailed).

Although the Part 1 and IELTS scores did not correlate to any great extent, Table 11 shows that the candidates passing Part 1 typically scored higher across the IELTS test than the failing candidates. Interestingly, candidates who failed Part 1 on first attempt gained a marginally higher mean score on the speaking component than the candidates who achieved a pass.

Table 11. IELTS scores by Part 1 outcome (first take, n=2093).

IELTS score	Part 1 result		Difference (Fail - Pass)	Std. Deviation	
	Fail	Pass		Fail	Pass
Reading	7.89	8.03	-0.14	.65	.68
Speaking	7.69	7.67	0.02	.58	.64
Understanding	8.17	8.24	-0.07	.56	.60
Writing	7.23	7.24	-0.01	.32	.39
Overall	7.82	7.87	-0.05	.33	.40

For completeness of reporting, Table 12 shows the IELTS scores by Part 1 outcome. As we would hope, this shows that the Part 1 fail rate reduces as the IELTS score rises, from 71.4% failing at an IELTS score of 7.0 (the pre-June 2014 threshold), to 6.7% at an IELTS score of 9.0.

⁶ For comparison, correlations for all candidates combined (all number of attempts) were: Reading=.107**, Speaking=-.024, Understanding=.058**, Writing=.036, Overall=.062**.

Table 12. Part 1 outcomes by IELTS score.

IELTS Overall Score		Part 1 outcome		Total
		Fail	Pass	
7.0	N	5	2	7
	%	71.4	28.6	100.0
7.5	N	369	876	1245
	%	29.6	70.4	100.0
8.0	N	359	956	1315
	%	27.3	72.7	100.0
8.5	N	80	256	336
	%	23.8	76.2	100.0
9.0	N	1	14	15
	%	6.7	93.3	100.0
Total	N	814	2104	2918
	%	27.9	72.1	100.0

Section 2: Part 1 item analysis

This section records the performance of the items used in Part 1 examinations in 2015-2016. Of the 800 items, 64.1% (n=345) were used once only, 24.3% twice (n=131), 10.2% three times (n=55) and 1.3% were used in all four tests in this period (n=7).

Table 13 shows the distribution of item facility and discrimination values by the three blueprint domains. While bearing in mind that the values represent the mean values across all items, tests and candidates, the relatively smaller group of items classified as *Applying knowledge and experience to clinical practice* appeared to be the most accessible items, with an average facility value of .70. The minimum values for facility show that at least one item within each domain was not answered correctly by any of the candidates, including the most able, and some produced a negative discrimination value. These items are highlighted in Appendix 1.

Table 13. Item facility and discrimination by blueprint domain

Item domain	N	FACILITY				DISCRIMINATION			
		Facility	SD	Min	Max	Discrim	SD	Min	Max
Good clinical care: assessment	360	0.67	0.21	0.00	0.98	.28	.11	-.10	.49
Good clinical care: management	294	0.66	0.20	0.00	0.97	.26	.13	-.27	.53
Applying knowledge and experience to clinical practice	146	0.70	0.20	0.00	0.97	.26	.14	-.20	.52

Table 14 shows the average item facility and discrimination values across 15 classifications of skill. Three of the skill areas were especially well represented across the Part 1 test in 2015-2016: *acute including emergency* (n=111), *epidemiology and health promotion* (n=190), and *investigation: interpretation of results* (n=169). The least sampled categories were *diagnosis* (n=7), *perioperative* (n=7) and *symptom relief and end of life* (n=8).

While bearing in mind that these values represent an average over multiple items, tests and candidates, the items relating to *genetics*, *perioperative*, and *social science and statistics*, on average, are among the most accessible. The most difficult items are the eight items focused on *symptom relief and end of life*. Interestingly, with the exception of the seven *perioperative* items, the *symptom relief and end of life* items produced the highest minimum score (.39), meaning that each of these items prompted at least 40% of correct responses (although some of this will be explained by the guessing factor for single best answer items).

Table 14. Item facility and discrimination by blueprint skill.

Item skill	N	FACILITY				DISCRIMINATION			
		Facility	SD	Min	Max	Discrim	SD	Min	Max
Acute including emergency	111	.68	.20	.10	.96	.25	.14	-.20	.52
Anatomy	32	.67	.18	.19	.95	.29	.11	.00	.47
Diagnosis	3	.66	.35	.28	.97	.26	.13	-.27	.53
Epidemiology and health promotion	190	.69	.20	.17	.98	.27	.13	-.10	.42
Genetics	21	.76	.18	.33	.96	.30	.09	.12	.47
Investigation: choice of appropriate tests	18	.73	.20	.09	.93	-	-	-	-
Investigation: interpretation of results	169	.65	.22	.00	.97	.26	.13	-.14	.49
Long-term	85	.66	.19	.07	.97	.25	.14	-.12	.52
Pathology and microbiology	25	.67	.15	.29	.93	.26	.09	.09	.42
Perioperative	7	.75	.12	.59	.94	.24	.10	.05	.45
Pharmacology and biochemistry	25	.65	.25	.09	.97	.30	.12	.00	.49
Physiology	28	.67	.26	.00	.91	.30	.11	.15	.49
Social science and statistics	23	.74	.17	.19	.96	.20	.13	.02	.33
Symptom relief and end of life	8	.59	.15	.39	.80	.29	.14	-.09	.50
Symptom	55	.63	.22	.00	.90	-	-	-	-

Table 15 shows the same data by blueprint topic area. The most accessible items were the 22 items on *urology* and 24 items on *developmental problems*. The most difficult items were the 34 *musculoskeletal* items and 33 *renal* items, which, on average, were answered correctly by just over half of all candidates.

Table 15. Item facility and discrimination by blueprint topic.

Item topic	N	FACILITY				DISCRIMINATION			
		Facility	SD	Min	Max	Discrim	SD	Min	Max
Blood and lymph	17	.73	.16	.19	.89	.30	.09	.11	.44
Breast	32	.64	.19	.21	.92	.32	.10	.07	.48
Cardiovascular	68	.64	.24	.10	.95	.26	.12	.00	.53
Developmental problems	24	.78	.18	.09	.96	.25	.06	.08	.34
Digestive	81	.64	.20	.00	.93	.24	.10	.00	.46
Endocrine	30	.66	.27	.00	.90	.27	.14	-.09	.46
ENT	33	.68	.22	.17	.97	.30	.12	.03	.53
Eye	12	.69	.22	.10	.95	.25	.13	.01	.44
Genitourinary	16	.64	.21	.23	.90	.29	.11	.10	.45
Homeostatic	33	.70	.22	.00	.91	.33	.10	.00	.46
Infectious disease	26	.67	.14	.33	.84	.32	.11	.12	.52
Mental health	100	.70	.19	.00	.98	.27	.13	-.20	.52
Musculoskeletal	34	.57	.21	.10	.88	.23	.15	-.12	.47
Neurological	60	.72	.16	.33	.97	.28	.10	.03	.46
Renal	33	.56	.25	.05	.95	.21	.16	-.16	.49
Reproductive	60	.67	.22	.15	.98	.23	.15	-.14	.49
Respiratory	45	.71	.15	.30	.97	.28	.12	-.05	.50
Seriously ill patient	50	.66	.19	.22	.94	.22	.16	-.27	.45
Skin	23	.66	.20	.18	.96	.28	.16	-.14	.52
Urological	22	.80	.15	.39	.96	.25	.11	.02	.43

The modified Angoff method is used to standard set Part 1. This method relies on judgement to determine the proportion of borderline candidates expected to answer an individual item correctly, bearing in mind the intended challenge and difficulty of the item. The standard setting panel's Angoff value can be conceived as a measure of *perceived* difficulty, as a low Angoff value is given where it is perceived that fewer borderline candidates will answer correctly (a difficult item), and a higher value is given where it is perceived that a higher proportion of borderline candidates will answer correctly (an easy item). As part of the test analysis it is prudent to check perceived difficulty (the Angoff value) against the observed difficulty (the facility value) as an indicator of the panel's efficacy in judging item difficulty for borderline candidates.

The next two tables cross-tabulate perceived item difficulty by observed item difficulty, highlighting where there are differences between the two indicators. The blue shading shows the number of items rated by the panel as less difficult than the candidates found them to be, while the red shading indicates the items that the panel rated as more difficult than the candidates found them to be. The emboldened cells indicate where the panel's judgement was consistent with the observed difficulty: in total, matching occurred for 28.0% of 'difficult' items, 45.7% of 'moderate' items and 32.0% of 'easy' items. In terms of the observed difficulty, in total, 5.3% of all items delivered in 2015-2016 proved difficult, 39.8% were of moderate difficulty, and 55.0% were found to be relatively easy for this cohort of candidates.

Table 16. Comparison of Angoff (perceived difficulty) by facility (observed difficulty)⁷.

		Angoff score (perceived difficulty)			Total
		Difficult <5.89	Moderate 5.90-6.73	Easy >6.74	
Facility (observed difficulty)	Difficult <.25	12	26	4	42
	Moderate .26-.69	120	145	52	317
	Easy >.70	71	227	140	438
Total		203	398	196	797*

*Missing Angoff values for three items.

Table 17. Comparison of Angoff (perceived difficulty) and facility (observed difficulty) by domain.

Item domain	Facility (observed difficulty)	Angoff (perceived difficulty)			Total
		Difficult <5.89	Moderate 5.90-6.73	Easy >6.74	
Applying knowledge and experience to clinical practice	Difficult <.25	5	2		7
	Moderate .26-.69	29	16	4	49
	Easy >.70	14	47	29	90
	Total	48	65	33	145
Good clinical care: assessment	Difficult <.25	2	16	2	20
	Moderate .26-.69	42	73	24	139
	Easy >.70	29	100	69	198
	Total	73	189	95	357
Good clinical care: management	Difficult <.25	6	8	2	16
	Moderate .26-.69	49	56	24	129
	Easy >.70	27	80	42	149
	Total	82	144	68	294

The item facility and discrimination values for all items used in 2015-2016 are presented in Appendix 1.

Section 3: Part 2 results

Section 3 explores total test scores, pass rates and the number of stations passed for the Part 2 OSCE by a number of test and candidate characteristics.

Section 3a: Part 2 results

Table 18 shows the pass rate for each OSCE administration and the Cronbach's alpha coefficient. Like Part 1, the pass rate shows variability across the different administrations, ranging from the lowest pass rate of 44.4% (n=45) to the highest pass rate of 95.8% (n=48). This variability is likely to be explained by fluctuations in cohort ability and station difficulty across circuits.

⁷ The Angoff values were banded using the quartile thresholds, so that 'difficult' items were defined as below the first quartile, 'moderate' items were the second and third quartiles and 'easy' was the fourth quartile. The observed difficulty used the same thresholds of <0.25, 0.26-0.75, and >.76.

As is usually the case due to the smaller number of assessment items, the OSCE tends to produce lower Cronbach's alpha coefficients than the Part 1 test; an average alpha of .657 compared with .915 for Part 1. The average alpha of 0.66 for Part 2 in 2015-2016 is consistent with that achieved in previous years. Some of the OSCE circuits produced alphas between 0.40 and .50, but these appear atypical and may be explained by cohort trends within individual circuits within each administration. Across all Part 2 administrations, 50% of the pass marks fell between 63% and 79%, while 50% of the alpha coefficients fell between .61 and .73.

Table 18. Part 2 pass/fail rate and Cronbach's alpha by diet⁸.

Part 2	Fail (%)	Pass (%)	Total (count)	Alpha
203682	37.20	62.80	43	.534
203683	36.40	63.60	44	.713
203684	34.10	65.90	41	.665
203685	28.30	71.70	46	.625
203686	25.00	75.00	32	.504
203687	26.20	73.80	42	.629
203688	23.40	76.60	47	.542
203689	21.70	78.30	46	.620
203690*	48.90	51.10	47	.738
203691	52.40	47.60	42	.728
203692	34.50	65.50	87	.611
203693	24.40	75.60	90	.458
203695	32.60	67.40	46	.599
203696*	27.10	72.90	48	.800
203697	41.90	58.10	43	.794
203698	23.40	76.60	47	.734
203699	34.00	66.00	47	.556
203843	47.90	52.10	48	.721
203844	55.60	44.40	45	.397
203845	47.80	52.20	46	.709
203846	36.20	63.80	47	.730
203847	15.60	84.40	45	.720
203848	30.20	69.80	43	.735
203849	24.40	75.60	41	.705
203850	50.00	50.00	46	.463
203851	10.40	89.60	48	.610
203852	39.50	60.50	43	.631
203853	23.40	76.60	47	.805
203854	17.40	82.60	46	.704
203855	51.20	48.80	41	.617
203856	15.60	84.40	45	.711
203857	53.50	46.50	43	.617
203858	12.50	87.50	40	.523
203863	9.10	90.90	44	.604
203864	18.60	81.40	43	.555
203865	27.90	72.10	43	.606
203886	32.60	67.40	43	.606
203887	27.90	72.10	43	.779
203902	23.30	76.70	43	.610
203903	27.70	72.30	47	.514
203964	21.30	78.70	47	.680
203965	38.30	61.70	47	.791
203966	18.00	82.00	50	.696
203967*	5.80	94.20	52	.703
203968	19.20	80.80	52	.614
203969	26.90	73.10	52	.645
203970	14.90	85.10	47	.678
203971	4.20	95.80	48	.639
203972	43.20	56.80	44	.709
203973	30.20	69.80	43	.786
204143	33.30	66.70	54	.796
204144	17.60	82.40	51	.750
204203	48.10	51.90	54	.762
204204	25.00	75.00	48	.642
Average	29.50	70.50	100.0	.657
Total	765	1826	2591	

Table 19 shows the Part 2 pass and fail outcomes for all IMG and UK candidates by the number of attempts at Part 2 in 2015-2016. Full attempt data are not available for Part 2 prior to 2015-2016, so the data below refer only to performance in this period, and not previous attempts that were taken prior to 2015. As such, where tables refer to first take candidate only, conclusions should be drawn with caution as the data may include a small proportion of retake candidates.

Like Part 1, the general trend for Part 2 is also towards a declining pass rate with each successive attempt at the OSCE. Interestingly, there is an upward trend after the third attempt, with a marginal increase in pass rate for the fourth attempt (+1.2%). The Part 2 mean score for failing candidates is remarkably stable, varying only between

⁸ Circuits 203690, 203696, and 203967 comprised 13 stations. Average scores have been re-calculated accordingly.

49.2% and 49.4%; it is not possible to determine whether this is a real trend or a product of reduced reliability in the attempt data (in that the number of the attempt is not wholly accurate and therefore not a true discriminator between candidates).

Table 2 showed that there was approximately 20% difference between the mean scores of the passing and failing candidates for Part 1, suggesting a reasonable spread of achieved scores. For Part 2, the difference between the passing and failing mean scores across all test administrations is 12.2%, which suggests some clustering of the cohort within a narrower range of scores. The explanations for this include that a) the raw score mark range is much smaller for Part 2 (maximum raw score of 54, compared with 200 for Part 1) meaning there is a less opportunity to spread the candidate scores; and b) Part 2 scores are an average (rather than a sum as for Part 1) across the 14 stations meaning that there is some compensation for performance across stations (e.g. with good performances cancelling out the effects of any weaker performances), and therefore some regression to the mean. Interestingly, the mean scores of the passing and failing candidates for Part 2 do not vary by the number of the attempt.

In Part 1 the proportion of IMG and UK candidates passing on first attempt was consistent during 2015-2016 at approximately 78/79%. However, as a group, IMG candidates were more likely to pass on first take than the UK candidates (79% and 72% respectively). For Part 2, the trends are similar, with 70/71% of UK and IMG candidates passing on first attempt, and, within region, IMG candidates slightly more likely than UK candidates to achieve a pass for Part 2 on first attempt (73.5% for IMG, compared with 68.7% among UK candidates). Across all attempts, the average pass rate of IMG and UK candidates is very similar (70.5% and 69.7% respectively).

Table 19. Part 2 pass rate by selected nationality world regions and number of attempts⁹.

Attempt		IMG			UK			Total			Average Part 2 score (%)	
		Fail	Pass	Total N / %	Fail	Pass	Total N / %	Fail	Pass	Total N / %	Fail	Pass
1 st	Result (%)	28.9	71.1	1692	30.1	69.9	163	29.0	71.0	1875	49.3	61.7
	In region (%)	71.5	73.5	72.9	68.1	68.7	68.5	71.0	72.9	72.4		
2 nd	Result (%)	29.2	70.8	452	23.4	76.6	47	28.9	71.1	506	49.4	60.9
	In region (%)	19.3	19.6	19.5	15.3	21.7	19.7	19.1	19.7	19.5		
3 rd	Result (%)	36.1	63.9	133	47.4	52.6	19	37.2	62.8	156	49.4	61.0
	In region (%)	7.0	5.2	5.7	12.5	6.0	8.0	7.6	5.4	6.0		
4 th	Result (%)	34.9	65.1	43	33.3	66.7	9	33.3	66.7	54	49.2	60.1
	In region (%)	2.2	1.7	1.9	4.2	3.6	3.8	2.4	2.0	2.1		
Total	Result (%)	29.5	70.5	2320	30.3	69.7	238	29.5	70.5	2591	49.5	61.7
	In region (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0		

Table 20 shows the descriptive statistics for each Part 2 administration. The blue shading highlights the circuits where the percentage score is more than one standard deviation below the mean for that particular variable, indicating lower than average scores. The red shading highlights circuits where the percentage score is more than one standard deviation above the mean, indicating higher than average scores. The total variability observed in the circuit-level pass mark is consistent with that of previous years, and is likely to be explained by minor fluctuations in cohort ability and the combination of station tasks. However, the shading suggests that there has been some degree of score creep in later administrations with more of the higher mean scores observed within the more recent administrations. This may be attributable to candidates being better prepared or having a better understanding of the Part 2 requirements.

⁹ Data are not shown for EEA candidates (n=20) and the Unspecified (n=13) category due to small numbers.

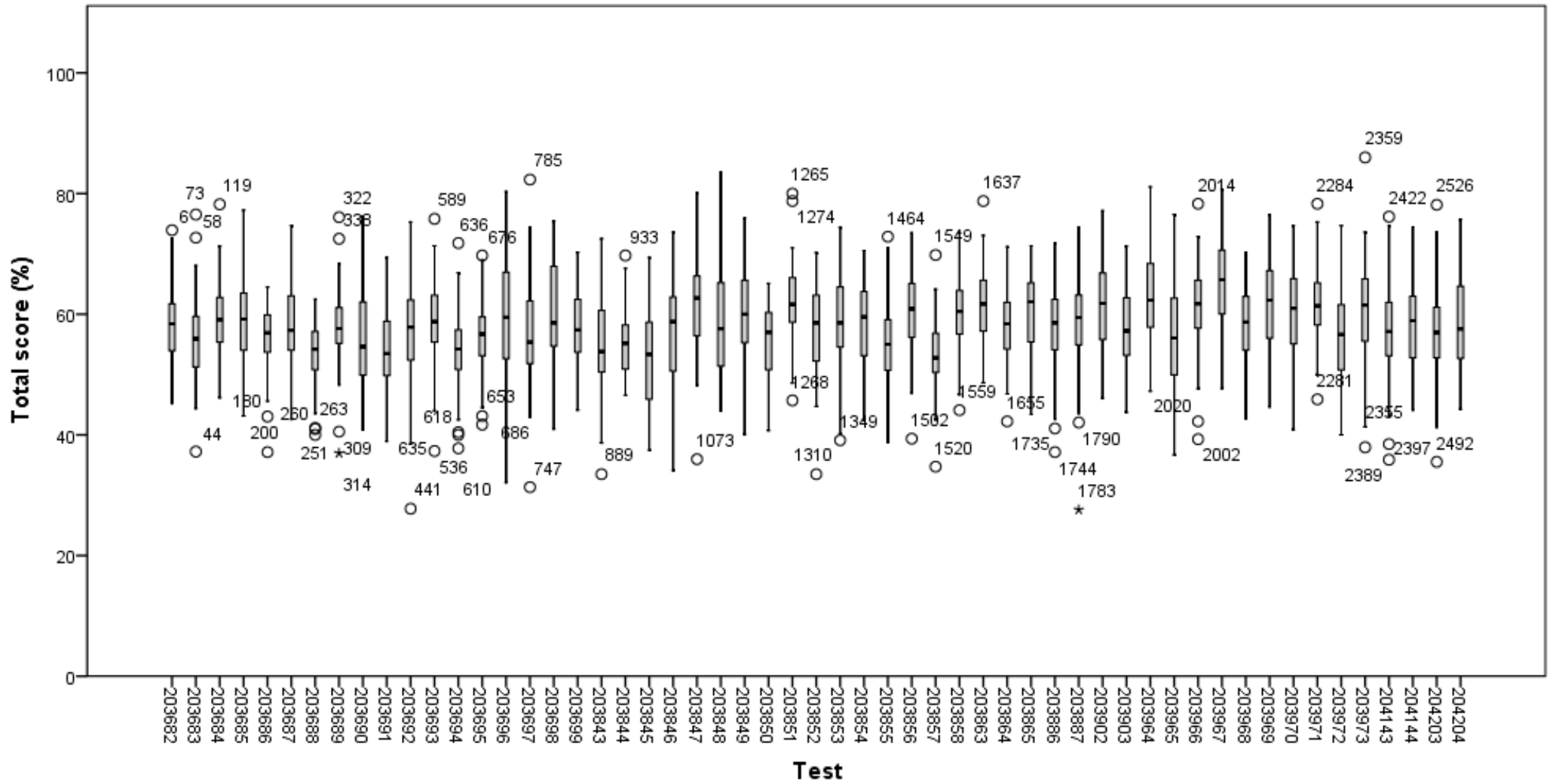
Table 20. Descriptive data for Part 2.

TESTID	Mean	95% Confidence Interval for Mean		5% Trimmed Mean	Median	Variance	SD	Min	Max	Range
		Lower Bound	Upper Bound							
203682	58.11	56.14	60.09	57.91	58.39	41.14	6.41	45.18	73.93	28.75
203683	55.77	53.52	58.01	55.56	55.94	54.41	7.38	37.23	76.52	39.29
203684	58.95	56.78	61.13	58.79	59.11	47.35	6.88	46.16	78.21	32.05
203685	58.99	56.98	61.00	58.91	59.20	45.66	6.76	43.13	77.23	34.11
203686	56.01	53.81	58.21	56.48	56.92	37.19	6.10	37.14	64.46	27.32
203687	57.99	55.90	60.08	57.98	57.32	45.09	6.71	42.59	74.64	32.05
203688	53.00	51.36	54.64	53.19	54.2	31.14	5.58	40.00	62.50	22.50
203689	58.12	56.05	60.18	58.27	57.63	48.55	6.97	36.96	76.07	39.11
203690	55.55	53.16	57.94	55.42	54.62	66.48	8.15	40.77	76.15	35.38
203691	54.20	51.91	56.49	54.17	53.48	54.13	7.36	38.93	69.38	30.45
203692	57.16	55.60	58.73	57.38	57.86	53.94	7.34	27.77	75.27	47.50
203693	58.73	57.38	60.07	58.88	58.75	41.42	6.44	37.32	75.80	38.48
203694	54.00	51.80	56.2	54.01	54.24	52.19	7.22	37.77	71.79	34.02
203695	56.07	54.24	57.91	56.10	56.7	38.18	6.18	41.70	69.73	28.04
203696	59.43	56.40	62.45	59.59	59.48	108.50	10.42	32.12	80.29	48.17
203697	56.56	53.86	59.26	56.44	55.36	77.09	8.78	31.34	82.32	50.98
203698	60.57	58.08	63.06	60.73	58.57	71.77	8.47	40.98	75.45	34.46
203699	57.68	55.78	59.58	57.73	57.41	41.76	6.46	44.11	70.27	26.16
203843	54.63	52.34	56.92	54.8	53.79	62.20	7.89	33.48	72.50	39.02
203844	55.27	53.52	57.01	55.01	55.18	33.69	5.80	46.61	69.73	23.13
203845	52.80	50.46	55.15	52.72	53.35	62.30	7.89	37.50	69.38	31.88
203846	56.80	54.20	59.40	57.13	58.75	78.47	8.86	34.11	73.57	39.46
203847	61.09	58.76	63.43	61.32	62.68	60.52	7.78	35.98	80.09	44.11
203848	58.81	56.04	61.58	58.53	57.59	80.87	8.99	44.02	83.57	39.55
203849	59.61	57.07	62.15	59.82	60.00	64.74	8.05	40.09	75.89	35.80
203850	55.61	53.84	57.38	55.86	57.01	35.42	5.95	40.71	65.09	24.38
203851	61.77	59.85	63.69	61.64	61.65	43.77	6.62	45.71	80.00	34.29
203852	57.58	55.21	59.95	57.96	58.57	59.34	7.70	33.48	70.18	36.70
203853	59.17	56.56	61.78	59.47	58.57	79.03	8.89	39.11	74.38	35.27
203854	58.30	56.18	60.42	58.55	59.55	50.95	7.14	42.41	70.54	28.13
203855	55.03	52.87	57.18	54.83	55.00	46.60	6.83	38.75	72.86	34.11
203856	60.52	58.33	62.71	60.76	60.89	53.22	7.30	39.38	73.48	34.11
203857	53.38	51.43	55.33	53.44	52.77	40.17	6.34	34.73	69.82	35.09
203858	60.18	58.15	62.21	60.33	60.45	40.20	6.34	44.11	73.48	29.38
203863	61.48	59.56	63.40	61.40	61.70	39.98	6.32	48.66	78.75	30.09
203864	58.11	56.27	59.95	58.23	58.39	35.75	5.98	42.23	71.16	28.93
203865	60.16	58.05	62.27	60.36	62.05	46.96	6.85	43.39	71.34	27.95
203886	57.69	55.54	59.84	58.06	58.57	48.78	6.98	37.14	71.79	34.64
203887	57.96	55.18	60.74	58.32	59.46	81.74	9.04	27.59	74.38	46.79
203902	61.47	59.18	63.77	61.55	61.79	55.82	7.47	46.07	77.14	31.07

TESTID	Mean	95% Confidence Interval for Mean		5% Trimmed Mean	Median	Variance	SD	Min	Max	Range
		Lower Bound	Upper Bound							
203903	57.81	55.95	59.68	57.8	57.23	40.39	6.36	43.75	71.25	27.50
203964	62.78	60.42	65.15	62.66	62.32	64.94	8.06	47.23	81.07	33.84
203965	56.08	53.44	58.72	56.08	56.07	80.76	8.99	36.70	76.52	39.82
203966	60.83	58.67	62.99	61.09	61.74	57.72	7.60	39.29	78.30	39.02
203967	65.02	62.85	67.19	65.12	65.72	60.87	7.80	47.69	80.67	32.98
203968	58.20	56.52	59.87	58.33	58.71	36.24	6.02	42.68	70.27	27.59
203969	61.22	59.20	63.23	61.32	62.32	52.47	7.24	44.64	76.52	31.88
203970	60.17	58.05	62.29	60.30	60.98	52.23	7.23	40.89	74.64	33.75
203971	61.44	59.6	63.27	61.36	61.38	39.94	6.32	45.89	78.30	32.41
203972	56.10	53.70	58.50	56.05	56.65	62.42	7.90	40.00	74.64	34.64
203973	60.40	57.59	63.21	60.49	61.52	83.38	9.13	37.95	85.98	48.04
204143	56.92	54.56	59.28	56.93	57.14	74.77	8.65	35.89	76.16	40.27
204144	58.18	56.12	60.23	58.10	58.93	53.22	7.30	44.11	74.46	30.36
204203	56.70	54.55	58.84	56.69	56.96	61.77	7.86	35.54	78.13	42.59
204204	58.79	56.63	60.94	58.66	57.54	55.01	7.42	44.29	75.71	31.43
Average	58.09	55.91	60.27	58.16	58.33	55.14	7.35	40.05	74.58	34.52
SD	2.59	2.59	2.64	2.62	2.78	15.90	1.04	4.88	4.61	6.47

Chart 2 shows that the variability in mean score across Part 2 administrations is within a narrow section of the total score range. Indeed, 50% of the mean circuit scores fall within a range of just 2%, between 58.1% and 60.2%.

Chart 2. The distribution of Part 2 scores by OSCE administration.



Section 3b: Part 2 performance by candidates' demographic characteristics

This section explores Part 2 performance by the following selected demographic characteristics of trainees within each test cohort: gender, age, ethnic origin, country of primary medical qualification, and nationality. The following should be noted:

- This section focuses on first take candidates only, and excludes from the analysis all re-take candidates. However, as previously noted, the data on number of takes may not be reliable and conclusions should be drawn with caution;
- The totals for each analysis may not be consistent as each is determined by the availability or completeness of the demographic profile of each candidate.

Table 21 shows the pass rate, average number of stations passed and descriptive scores for Part 2 for male and female candidates. This shows that male candidates achieved a marginally higher pass rate (+1.4%) than female candidates but, in all other respects, there are very few differences in performance: zero difference in terms of the average number of stations passed and a difference of 0.1% in the mean score. Further analysis is presented below to show that there is a statistically significant difference in Part 2 score by gender. However, those differences are noted to be marginal and are not apparent when referring to overall cohort trends across multiple administrations.

Table 21. Part 2 outcome, number of station passes, and total score by candidates' gender (first take).

Part 2			Gender		Total/ Average
			Male	Female	
Result	Fail	% result	44.2	55.8	543
		% gender	28.1	29.5	
	Pass	% result	45.9	54.1	1308
		% gender	71.9	70.5	
Stations passed (count)	Mean		9.8	9.8	9.8
	Std. Deviation		2.3	2.4	2.38
Score (%)	Mean		58.3	58.4	58.4
	SD		7.7	8.1	7.9
	Minimum		27.8	31.3	29.6
	Maximum		83.6	86.0	84.8
	Range		55.8	54.6	55.2
	IQR		9.9	10.6	10.3
Average			841	1010	1851

Table 22 shows the same data by candidates' age. Like for Part 1, candidates aged 32 and under tended, on average, to gain the highest scores, pass rates and average number of stations passed. The trend that was noted for Part 1 among candidates aged 52 or over is not replicated for Part 2.

Table 22. Part 2 outcome, number of station passes, and total score by candidates' age (first take).

Part 2			Age						Total/ Average
			21-26	27-32	33-38	39-44	45-51	52+	
Result	Fail	% result	17.6	49.7	20.7	7.9	3.4	0.7	100.0
		% age	27.1	27.3	31.7	40.4	42.9	33.3	543
	Pass	% result	19.6	54.5	18.5	5.2	1.7	0.6	100.0
		% age	72.9	72.7	68.3	59.6	57.1	66.7	1308
Stations passed (count)	Mean		10.1	9.9	9.6	9.1	8.9	9.6	9.5
	SD		2.4	2.4	2.4	2.3	2.9	3.1	2.6
Score (%)	Mean		59.1	58.7	57.7	56.5	55.6	57.5	57.5
	SD		7.9	7.8	7.8	8.1	9.6	8.6	8.3
	Min		27.8	33.5	35.9	35.5	31.3	38.5	33.8
	Max		83.6	86.0	82.3	77.8	78.1	70.3	79.7
	Range		55.8	52.5	46.4	42.2	46.8	31.8	45.9
	IQR		10.1	10.0	9.6	10.6	13.2	11.6	10.8
Total			352	984	355	110	40	12	1853

Table 23 shows the Part 2 descriptive data by candidates' ethnic origin. The largest group of candidates are of Asian or Asian British ethnic origin, and comprised 55.3% of the first take candidature in 2015-2016. The small groups of candidates who classified themselves as White, or did not state their ethnic origin ('Not stated' in the table below) comprised 6.5% and 4.3% respectively of the first take candidature and, on average, these two groups outperformed all other ethnic groups based on pass rate, mean score and average number of stations passed. As is discussed further below, the observed differences in the scores achieved by each ethnic group were not found to be statistically significant and, as such, are attributed to chance occurrences alone.

Table 23. Part 2 outcome, number of station passes, and total score by candidates' ethnic origin (first take).

Part 2			Ethnic origin						
			Asian or Asian British	Black or Black British	Mixed	Not stated	Other Ethnic Groups	Unspecified	White
Result	Fail	% result	54.3	21.7	1.8	3.5	10.5	2.8	5.3
		% age	28.8	32.4	33.3	23.8	29.5	38.5	24.0
	Pass	% result	55.7	18.8	1.5	4.7	10.4	1.8	7.0
		% age	71.2	67.6	66.7	76.3	70.5	61.5	76.0
Stations passed	Mean		9.78	9.6	9.8	10.2	9.8	9.7	10.2
	SD		2.3	2.5	2.8	2.3	2.2	2.4	2.7
Score	Mean		58.2	58.1	58.6	59.6	57.9	57.6	60.2
	SD		7.5	8.2	9.8	8.3	7.3	7.5	10.2
	Min		34.7	33.5	39.4	40.7	27.8	44.0	31.3
	Max		80.7	86.0	78.8	78.2	82.3	72.2	81.1
	Range		45.9	52.5	39.4	37.5	54.6	28.2	49.7
	IQR		9.9	10.9	14.1	11.5	9.1	10.4	13.7
Total			1035	368	30	80	189	31	120

Table 24 shows pass rate and mean score by candidates' country of primary medical qualification. The highest performing group for Part 1 comprised the candidates qualifying in Myanmar, with a pass rate of 97.4%. For Part 2, the candidates qualifying in Myanmar achieved a pass rate of 71%, and were outperformed by candidates qualifying in a number of other countries. The most consistently high performing group across Parts 1 and 2 comprised the candidates who qualified in Australia (average pass rate Part 1=97.0% and Part 2=89.7%).

Among the national cohorts comprising ten or more candidates, the candidates qualifying in Sri Lanka underperformed relative to all other sizeable cohorts, with a pass rate that was 17.4% below the average (53.3% compared to the average of 71.1%). However, the average number of stations passed (data not shown) and the mean score for those qualifying in Sri Lanka were close to the mean, suggesting that, on average, these candidates are falling just short of the within-station standards required to achieve a pass for Part 2.

Table 24. Pass rate and mean scores by country of primary medical qualification (first take)*.

PMQ	RESULT		N	Part 2 score	
	Fail	Pass		Mean	SD
Afghanistan	20.0	80.0	5	57.8	7.6
Algeria	█	█	█	█	█
Antigua And Barbuda	█	█	█	█	█
Argentina	40.0	60.0	5	57.3	6.2
Australia	10.3	89.7	39	65.9	10.2
Azerbaijan	█	█	█	█	█
Bahrain	15.8	84.2	19	62.3	7.2
Bangladesh	36.4	63.6	55	55.3	7.6
Barbados	█	█	█	█	█
Belarus	28.6	71.4	7	56.0	9.2
Bosnia And Herzegovina	█	█	█		
Brazil	28.6	71.4	7	57.5	9.3
Cameroon	█	█	█	█	█
Canada	█	█	█	█	█
Cayman Islands	█	█	█	█	█
Chile	█	█	█		
China	32.6	67.4	43	56.5	9.2
Colombia	█	█	█	█	█
Cuba	28.6	71.4	7	55.8	4.3
Czech Republic	0.0	100.0	4	58.5	3.6
Egypt	23.4	76.6	94	57.7	6.1
Ethiopia	█	█	█		
Fiji	█	█	█		
Gambia	█	█	█		
Georgia	█	█	█	█	█

	RESULT		N	Part 2 score	
	Fail	Pass		Mean	SD
Kuwait	█	█	█		
Kyrgyzstan	█	█	█		
Lebanon	█	█	█	█	█
Libya	33.3	66.7	12	59.5	10.3
Lithuania	█	█	█	█	█
Malaysia	0.0	100.0	5	62.1	1.8
Mauritius	█	█	█	█	█
Mexico	25.0	75.0	4	60.3	5.9
Moldova	█	█	█	█	█
Mongolia	█	█	█		
Myanmar	28.6	71.4	35	60.7	6.0
Nepal	41.7	58.3	24	56.6	7.7
New Zealand	40.0	60.0	5	56.4	11.0
Nigeria	34.0	66.0	247	56.9	7.8
Oman	█	█	█	█	█
Pakistan	29.0	71.0	514	57.9	7.4
Peru	█	█	█		
Philippines	22.2	77.8	9	56.9	7.9
Poland	42.9	57.1	7	56.9	9.1
Romania	40.0	60.0	5	59.8	7.5
Russia	28.1	71.9	32	58.2	6.8
Saba	█	█	█		
Saint Kitts And Nevis	20.0	80.0	5	60.6	6.1
Saudi Arabia	█	█	█	█	█
Serbia	75.0	25.0	4	52.7	15.0

PMQ	RESULT		N	Part 2 score	
	Fail	Pass		Mean	SD
Ghana	0.0	100.0	12	67.0	6.5
Grenada	20.0	80.0	10	60.6	8.1
Guinea	█	█	█		
Guyana	█	█	█		
Hungary	30.0	70.0	10	61.3	9.1
India	26.6	73.4	271	58.5	7.4
Iran	30.8	69.2	13	55.7	7.2
Iraq	25.6	74.4	39	58.4	6.6
Ireland	14.3	85.7	7	58.2	7.4
Israel	20.0	80.0	5	60.7	6.8
Italy	█	█	█		
Jamaica	22.2	77.8	9	59.4	5.5
Japan	█	█	█		
Jordan	22.2	77.8	9	59.9	5.4
Kazakhstan	50.0	50.0	4	53.7	9.2
Kenya	33.3	66.7	9	56.3	8.6
Korea	█	█	█		

	RESULT		N	Part 2 score	
	Fail	Pass		Mean	SD
South Africa	19.2	80.8	26	64.0	10.8
Sri Lanka	46.7	53.3	15	55.4	6.8
Sudan	37.2	62.8	78	57.3	9.0
Syrian Arab Rep	28.6	71.4	7	58.3	6.5
Taiwan	█	█	█	█	█
Tanzania	25.0	75.0	4	59.3	10.4
Thailand	█	█	█		
Turkey	28.6	71.4	7	56.6	12.7
Uganda	█	█	█	█	█
Ukraine	33.3	66.7	24	58.4	6.9
UAE	21.7	78.3	23	59.6	8.1
US	28.6	71.4	7	55.7	6.6
Uzbekistan	█	█	█	█	█
Venezuela	█	█	█	█	█
Zimbabwe	25.0	75.0	8	59.1	7.0
Average	28.9	71.1	1853	58.4	7.9

*Excludes data for 67 candidates who completed a 13-station circuit.

Table 25 shows the same data by candidates' nationality. Excluding the nationalities with fewer than 10 candidates (where some candidates have performed well), the Australian candidates clearly outperform all other sizeable national cohorts in terms of pass rate (+23.4% of the average), number of stations passed (nearly two stations more than average [n=+1.76]) and mean score (+7.45%). Among the larger national cohorts, the Bangladeshi candidates appear to underperform relative to all others (-11.7% of the mean pass rate; nearly a whole station below average [-0.97]; and -2.73% of the mean score).

Table 25. Pass rate and mean scores by nationality*.

Nationality	Part 2 result		N	Part 2 score	
	Fail	Pass		Mean	SD
Afghan	█	█	█	█	█
Albanian	█	█	█		
Algerian	█	█	█		
American	22.7	77.3	22	58.8	7.4
Antiguan	█	█	█		
Argentinian	40.0	60.0	5	57.3	6.2
Australian	5.9	94.1	34	66.1	9.0
Azerbaijani	█	█	█		
Bahamian	█	█	█	█	█
Bahraini	█	█	█	█	█

Nationality	Part 2 result		N	Part 2 score	
	Fail	Pass		Mean	SD
Korean	█	█	█		
Lebanese	█	█	█		
Libyan	42.9	57.1	7	60.5	12.3
Malaysian	23.5	76.5	17	55.7	4.3
Maldivian	█	█	█	█	█
Mauritian	25.0	75.0	8	59.1	8.7
Mexican	█	█	█	█	█
Mongolian	█	█	█		
Mozambican	█	█	█		
Myanmar	25.0	75.0	28	61.1	5.2

Nationality	Part 2 result		N	Part 2 score	
	Fail	Pass		Mean	SD
Bangladeshi	39.5	60.5	38	55.7	7.5
Barbadian					
Belarusian					
Bosnian					
Botswanian					
Brazilian	28.6	71.4	7	56.9	9.4
British	30.6	69.4	160	57.7	8.7
Burmese	50.0	50.0	6	58.3	9.8
Cameroonian					
Canadian	24.3	75.7	37	60.8	7.6
Chilean					
Chinese	20.0	80.0	5	57.1	8.7
Colombian					
Cuban					
Dominican					
Egyptian	24.4	75.6	86	58.0	6.7
Ethiopian					
Filipino	20.0	80.0	5	57.8	9.4
Finnish					
Gambian					
German					
Ghanaian	16.7	83.3	6	60.4	8.0
Grenadian					
Guyanese					
Hong Kong					
Indian	25.9	74.1	278	58.5	7.1
Iranian	23.1	76.9	13	56.1	6.9
Iraqi	24.4	75.6	41	59.1	6.8
Irish					
Israeli					
Ivorian					
Jamaican	16.7	83.3	6	60.4	5.9
Japanese					
Jordanian	20.0	80.0	10	60.8	5.6
Kenyan	40.0	60.0	10	55.2	8.8

Nationality	Part 2 result		N	Part 2 score	
	Fail	Pass		Mean	SD
Namibian					
Nepalese	33.3	66.7	33	57.1	8.5
New Zealander	0.0	100.0	4	64.6	5.9
Nigerian	31.6	68.4	266	57.6	7.9
Norwegian					
Omani					
Pakistani	29.0	71.0	486	57.8	7.5
Palestinian					
Peruvian					
Polish					
Portuguese					
Romanian					
Russian	27.3	72.7	11	59.7	8.2
Saudi Arabian					
Serbian	75.0	25.0	4	52.7	15.0
Singaporean	0.0	100.0	4	60.8	5.0
Somali					
South African	28.6	71.4	21	63.9	10.4
Sri Lankan	40.7	59.3	27	57.2	7.2
Sudanese	39.3	60.7	56	56.0	8.9
Swedish					
Syrian	50.0	50.0	6	55.4	3.3
Taiwanese	50.0	50.0	4	54.0	12.3
Tanzanian					
Thai					
Trinidad	33.3	66.7	6	60.8	3.4
Turkish	33.3	66.7	6	54.5	12.6
Ugandan					
Ukrainian					
Unspecified					
Uzbekistani					
Venezuelan					
Zambian					
Zimbabwean	22.2	77.8	9	58.6	6.7
Average	28.9	71.1	1853	58.4	7.9

*Excludes data for 67 candidates who completed a 13-station circuit.

Univariate analysis of variance was used to determine the effect of age, gender, and ethnic origin on the distribution of Part 2 scores. The results suggested that the combination of the three variables accounted for 18% of the observed score variance (less than that observed for Part 1 at 21%). Like Part 1, the candidates' age was statistically significant ($p=.003$). For Part 2 gender also produced a statistically significant result ($p=.009$). In both cases, the effect sizes were negligible meaning that although the trends could not be attributed to chance alone, candidates' age and gender

appear to have no real effect on their scores and outcomes ($\text{Eta}^2=.01$ and $.004$ respectively). None of the interactions that are shown in Table 26 below, between age, gender and ethnic origin, were significant.

Following on from the analysis of scores, the Chi-square test was used to explore if the Part 2 pass and fail results were distributed significantly differently from those that would be expected, all things being equal, across the demographic characteristics noted above. The findings suggested that there was no significant difference between the passing and failing candidates in terms of their gender. For age, the difference was statistically significant ($p=.027$). As was also noted for Part 1, the differences were mostly attributable to the age group 27-32 which achieved fewer fails (-6.8%) and more passes (+2.5%) than would be expected if all other things were equal. For ethnic origin, the differences were not statistically significant, and all differences between the observed and expected values were marginal.

Table 26 shows the mean score and pass rate for Part 2 by each of the demographic characteristics included in the ANOVA and chi-square analyses, and their interactions. As for the Part 1 analysis, splitting cohorts by the various interactions creates small groups of candidates and any conclusions should be drawn with caution.

Table 26. Part 2 scores and pass rates for interactions between age, gender and ethnicity.

Part 2		Pass rate (%)		Score			Total		
		Female	Male	Female	Male	Average	Female	Male	Total
21-26	Asian or Asian British	70.9	69.5	59.3	58.2	58.8	119	96	215
	Black or Black British	73.9	68.2	59.8	60.4	60.1	23	23	46
	Mixed	█	█	█	█	█	█	█	█
	Not stated	100.0	83.3	64.2	58.0	61.9	10	6	16
	Other Ethnic Groups	84.2	76.7	60.6	57.0	58.4	19	31	50
	Unspecified	100.0		65.2	47.4	58.1	3	2	5
	White	75.0	75.0	61.6	59.5	60.5	8	8	16
	Total	75.3	70.3	59.9	58.2	59.1	184	168	352
27-32	Asian or Asian British	68.8	78.5	57.8	58.9	58.2	375	229	604
	Black or Black British	73.5	67.0	60.2	58.2	59.0	71	99	170
	Mixed	75.0	80.0	61.9	60.1	61.3	12	5	17
	Not stated	66.7	72.7	58.3	59.7	59.0	21	22	43
	Other Ethnic Groups	72.1	70.6	57.8	58.6	58.2	43	50	93
	Unspecified		100.0		66.8	66.8		1	1
	White	82.4	90.9	61.4	64.4	62.6	34	22	56
	Total	70.6	75.4	58.4	59.0	58.7	556	428	984
33-38	Asian or Asian British	66.7	80.4	57.4	59.3	58.2	86	56	142
	Black or Black British	64.9	70.1	57.2	56.9	57.1	57	65	122
	Mixed	█		█	█	█	█	█	█
	Not stated	88.9	75.0	62.1	54.4	59.7	9	4	13
	Other Ethnic Groups	53.8	60.0	55.7	56.8	56.3	13	19	32
	Unspecified	37.5	60.0	56.0	56.7	56.3	6	4	10
	White	62.5	84.6	58.9	60.1	59.4	16	12	28
	Total	64.5	72.9	57.5	57.9	57.7	194	161	355
39-44	Asian or Asian British	61.3	59.1	56.9	57.6	57.2	32	22	54
	Black or Black British	66.7	57.1	58.2	54.0	55.6	9	14	23
	Not stated	100.0		62.8	47.5	57.7	4	2	6
	Other Ethnic Groups	50.0	60.0	54.7	56.5	56.1	2	7	9

Part 2	Pass rate (%)		Score			Total			
	Female	Male	Female	Male	Average	Female	Male	Total	
Part 2	Unspecified	75.0	40.0	55.9	55.3	55.6	4	3	7
	White	60.0	50.0	51.0	58.7	55.2	5	6	11
	Total	65.5	54.2	56.9	56.1	56.5	56	54	110
45-51	Asian or Asian British	70.0	33.3	55.6	56.8	56.1	10	7	17
	Black or Black British		16.7		51.6	51.6		6	6
	Mixed								
	Not stated			53.0		53.0	1		1
	Other Ethnic Groups	100.0	100.0	58.8	60.0	59.2	2	1	3
	Unspecified	100.0	75.0	63.9	58.7	60.8	2	3	5
	White								
	Total	73.9	36.8	56.6	54.5	55.6	21	19	40
52+	Asian or Asian British								
	Black or Black British								
	Not stated		100.0		64.1	64.1		1	1
	Other Ethnic Groups		50.0		60.4	60.4		2	2
	Unspecified	66.7		57.7		57.7	3		3
	White								
	Total	66.7	66.7	55.9	59.0	57.5	6	6	12
Total	70.0	71.5	58.4	58.3	58.4	1017	836	1853	

Section 2c. Part 2 other analyses, including IELTS scores

Table 27 shows the proportion of Part 2 passes and fails by candidates' overall IELTS score. The table shows that the majority of candidates scored between 7.5 - 8.5, and that those scoring 8.5 were the most likely to achieve a pass for Part 2. Although it appears that the candidates who achieved the highest score of 9 were just as likely as the lowest scoring IELTS candidates to achieve a pass in Part 2, the small numbers in both groups means this is unlikely to be a reliable finding.

Table 27. IELTS scores and Part 2 outcomes (first take).

IELTS overall score		Part 2 outcome		Total
		Fail	Pass	
7.0	Count	1	2	3
	%	33.3	66.7	100.0
7.5	Count	168	336	504
	%	33.3	66.7	100.0
8.0	Count	133	367	500
	%	26.6	73.4	100.0
8.5	Count	25	101	126
	%	19.8	80.2	100.0
9.0	Count	2	4	6
	%	33.3	66.7	100.0
Total	Count	329	810	1139
	%	28.9	71.1	100.0

Table 28 shows the Pearson correlation coefficients for the Part 2 scores with all IELTS component scores. Like for Part 1, the coefficients suggest statistically significant but weak correlations. As we might expect, after the overall IELTS score, the strongest relationship for Part 2 is with the scores achieved on the speaking component.

Table 28. Correlations between Part 2 total score and IELTSs component scores (first take).

Pearson correlation	Read	Speak	Understand	Write	IELTS Overall
Part2 score	.116**	.123**	.097**	.073**	.152**
Read		.245**	.454**	.310**	.716**
Speak			.317**	.409**	.653**
Understand				.318**	.706**
Write					.623**

** . Correlation is significant at the 0.01 level (2-tailed).

For the candidates taking both Part 1 and Part 2 in 2015-2016, the correlation between their Part 1 and Part 2 scores is 0.237 and is significant at the .001 level. For first take candidates only the correlation is 0.221 and significant at the .001 level.

Section 4: Part 2 station analysis

This section records the performance of each station used in the Part 2 OSCE examinations in 2015-2016. During this period 160 stations were used, and were administered and examined a total of 38,000 times.

Table 29 shows the descriptive statistics for the items used in this period by the four blueprint domains. While bearing in mind that these are average values across all stations, circuits and candidates, the *practical* skills stations appeared to gain the highest scores, while the *communication* skills stations, on average, gain the lowest scores.

Table 29. Average station scores by blueprint skill area.

Station skill	Facility					Discrimination
	N	Mean	SD	Min	Max	
Communication	57	0.56	0.17	0.00	1.00	0.41
Examination	29	0.58	0.19	0.00	1.00	0.42
History	65	0.59	0.17	0.00	1.00	0.40
Practical	9	0.60	0.23	0.00	1.00	0.41

Table 30 shows the average proportion of station passes and fails by blueprint skill area. Interestingly, while the table above shows that the *practical* skills stations, on average, produce the highest scores (albeit marginally), the table below shows that they also produce the lowest number of station passes (66.9%). This may be attributable to the higher pass mark that is observed across these stations (50.8% compared to 46.5% for *communication* skills; or raw scores of 2.03 and 1.86), which may reflect their focus on essential or core clinical skills.

Table 30. The proportion of station passes and fails by station skill area.

		SKILL AREA				Total*
		Comm.	Exam.	Hist.	Pract.	
Fail	N uses*	3199	3390	2951	1758	11298
	% result	28.3	30.0	26.1	15.6	100.0
	% skill	29.3	31.4	27.1	33.1	29.8
Pass	N uses*	7705	7414	7953	3553	26625
	% result	28.9	27.8	29.9	13.3	100.0
	% skill	70.7	68.6	72.9	66.9	70.2
Total	N uses*	10904	10804	10904	5311	37923
	% result	28.8	28.5	28.8	14.0	100.0
	% skill	100.0	100.0	100.0	100.0	100.0

*N and Total refer to total number of uses, not individual stations.

The full list of stations, station titles, and the mean scores and pass rates are presented in Appendix 2.

Section 5: Part 2 examiners

This section explores the performance of the 159 examiners who were active for Part 2 during 2015-2016. Of these, 130 examiners (81.8%) provided basic demographic data relating to their age, gender, and ethnic origin¹⁰. The following tables split the examiner cohort by these demographic characteristics (age, gender and ethnic origin) and report the number of candidates assessed in 2015-2016, the mean station scores awarded, the proportion of station passes and fails awarded, and the proportion of global judgements awarded (Borderline, Fail and Pass). Appendix 3 records the same data for each individual examiner.

Table 31 compares the Part 2 outcomes awarded by examiners by their gender. It is firstly notable that for every ten Part 2 examiners active in 2015-2016, approximately eight were male and two were female. The gender balance of examiners does not mirror that of the candidates, as, in 2015-2016, there were slightly more females than males attempting Part 1 (52.5% female) and Part 2 (54.6% female). On average, the female examiners tended to examine fewer candidates than the male examiners (an average of 237 candidates for each male examiner, and 187 candidates for each female examiner). Despite this imbalance in gender and the average number of candidates examined, the performance of male and female examiners seems comparable; the mean scores, proportion of passes and fails, and the global judgements awarded are all similar. The biggest observed difference is in relation to the use of the Borderline global judgement, with female examiners favouring this marginally more frequently than male examiners.

Table 31. Examiner performance by gender.

		Male	Female
N / %		101/77.7	29/22.3
Number of candidates	Mean	237	187
	SD	185	167
Mean station score	Mean	2.4	2.3
	SD	0.3	0.2
Fail results (%)	Mean	28.8	29.4
	SD	12.6	13.1
	Min	6.3	6.8
	Max	61.2	61.0
Pass results (%)	Mean	71.2	70.6
	SD	12.6	13.1
	Min	38.8	39.0
	Max	93.8	93.2
Borderline global judgements (%)	Mean	21.4	26.5
	SD	9.4	11.1
Fail global judgements (%)	Mean	17.5	17.2
	SD	9.5	9.2
Pass global judgements (%)	Mean	61.1	56.4
	SD	13.8	14.5

¹⁰ Data on religion, sexual orientation and disability were also collected but are not referred to in this report.

Table 32 shows the Part 2 outcomes by examiners' age (by pre-specified age band). The number of examiners aged 34 or under is small (4) and the data are provided for completeness only. The largest group comprises examiners aged between 55-64 (31.5%), while the group aged 65 or over were the most active in 2015-2016, based on the average number of candidates examined.

The examiner groups at the extremes of the age range appear marginally more likely to award slightly lower scores, which in turn produces slightly fewer pass results among these examiners. The greatest difference appears to be between those aged 35-44 (most likely to award pass outcomes) and those aged 65 or over (least likely to award pass outcomes). The outcomes of the examiners aged between 45-54 are unremarkable and consistent with those of the other mid-range groups. However, this group shows the largest disparity between the pass results and pass global judgements (-15.1%), and a slightly greater reliance on the Borderline global judgement, than the Pass global judgement for passing candidates. Statistical analysis has not been conducted to explore whether the trends observed in Table 32 are due to chance occurrences, or real effects on the basis of examiner age. Any trends should therefore be treated with caution.

Table 32. Examiner performance by age band.

		34 or under	35-44	45-54	55-64	65+
N / %		4/3.1	32/24.6	26/20.0	41/31.5	27/20.8
Number of candidates	Mean	104	118	157	254	396
	SD	68	76	147	170	196
Mean station score	Mean	2.2	2.4	2.4	2.4	2.3
	SD	0.3	0.2	0.3	0.3	0.2
Fail results (%)	Mean	30.3	26.8	28.0	28.9	32.4
	SD	18.1	12.3	15.4	12.5	9.7
	Min	6.8	6.8	6.3	10.4	12.7
	Max	50.0	49.5	61.2	61.0	49.1
Pass results (%)	Mean	69.7	73.2	72.0	71.1	67.6
	SD	18.1	12.3	15.4	12.5	9.7
	Min	50.0	50.5	38.8	39.0	50.9
	Max	93.2	93.2	93.8	89.6	87.3
Borderline global judgements (%)	Mean	23.3	24.6	26.1	20.0	20.5
	SD	14.3	7.9	12.3	10.3	7.5
Fail global judgements (%)	Mean	17.6	15.0	17.0	17.2	21.2
	SD	9.2	8.5	11.4	8.4	9.3
Pass global judgements (%)	Mean	59.1	60.5	56.9	62.8	58.3
	SD	7.0	10.3	17.8	14.8	13.3

Table 33 shows the same data by examiners' ethnic origin. The majority of examiners active in 2015-2016 were of Asian or Asian British ethnic origin (51.5%). The next largest group was White examiners (35.4%), with very few examiners with Black or Black British, Mixed or Other ethnic origins. Given the small numbers in some of these categories it is inadvisable to draw conclusions about any marking tendencies. Indeed, the table suggest that examiners who classified themselves as being of 'Other' ethnic origin appear to be more likely to award passing scores than the other ethnic groups, but this is based on the trends across five examiners only.

Table 33. Examiner performance by ethnic origin.

		Asian or Asian British	Black or Black British	Mixed	Other Ethnic Group	White
N / %		67/51.5	8/6.2	4/3.1	5/3.1	46/35.4
Number of candidates	Mean	189	226	233	77	295
	SD	177	144	183	65	184
Mean station score	Mean	2.3	2.4	2.4	2.5	2.3
	SD	0.3	0.4	0.2	0.2	0.2
Fail results (%)	Mean	29.4	26.0	24.0	11.8	30.8
	SD	12.9	16.1	10.1	7.9	11.5
	Min	6.3	9.1	14.6	6.8	9.1
	Max	61.0	61.2	36.4	23.4	60.9
Pass results (%)	Mean	70.6	74.0	76.0	88.2	69.2
	SD	12.9	16.1	10.1	7.9	11.5
	Min	39.0	38.8	63.6	76.6	39.1
	Max	93.8	90.9	85.4	93.2	90.9
Borderline global judgements (%)	Mean	23.9	21.8	19.7	27.3	20.7
	SD	11.0	12.2	5.3	11.8	7.9
Fail global judgements (%)	Mean	17.4	17.8	8.8	12.1	18.8
	SD	9.8	9.0	2.2	15.0	8.6
Pass global judgements (%)	Mean	58.7	60.4	71.5	60.7	60.5
	SD	15.8	15.7	5.2	11.8	11.4

Table 34 shows Part 2 outcomes by the secondary form of ethnic origin data that is requested of examiners; their national/geographical origin, which is shown below for the three largest groups only¹¹. The table shows that examiners of Indian origin are marginally more likely to award higher scores, which goes hand in hand with a greater proportion of pass outcomes and pass global judgements. Although the group of examiners identifying themselves as being of Pakistani origin is much smaller (14), there is a difference of 6.5% between the proportion of passing station results, as compared to those of the examiners of Indian origin.

¹¹ Table 34 excludes 26 examiners who were distributed in very small groups such as Caribbean (1), African (7), Bangladeshi (1).

Table 34. Examiner performance by national/geographical origin (selected groups only).

		Indian	British	Pakistani
N / %		46/35.4	44/33.8	14/10.8
Number of candidates	Mean	174	279	175
	SD	151	178	163
Mean station score	Mean	2.4	2.3	2.3
	SD	0.3	0.2	0.3
Fail results (%)	Mean	28.0	31.1	34.5
	SD	12.6	11.3	14.0
	Min	6.3	9.1	13.6
	Max	57.4	60.9	61.0
Pass results (%)	Mean	72.0	68.9	65.5
	SD	12.6	11.3	14.0
	Min	42.6	39.1	39.0
	Max	93.8	90.9	86.4
Borderline global judgements (%)	Mean	23.6	21.2	25.1
	SD	10.0	8.1	11.9
Fail global judgements (%)	Mean	16.7	18.9	19.4
	SD	10.6	8.5	8.8
Pass global judgements (%)	Mean	59.7	59.9	55.5
	SD	15.5	11.3	16.0

CONCLUDING COMMENTS

This report has explored the performance of PLAB Parts 1 and 2 for 2015-2016 over a number of test facets, including candidates, test administrations, items/stations and examiners. A wide range of classification data is recorded for candidates and for the PLAB assessment items, and these were also used to check for trends at a cohort- or within-classification level. This analysis allows us to understand the extent to which the examinations are performing as intended, and producing reliable and defensible outcomes across the diversity of the candidature.

When combining all candidates and administrations, overall, both Part 1 and 2 produced fairly high pass rates (72.3% and 70.5% respectively). That said, for every administration of Part 1, approximately three out of every ten candidates were taking their examination for at least the second time, with a small proportion making multiple attempts, and with a decreasing likelihood of success at each attempt. It was found that candidates' age explained some of the variance observed in Part 1 and Part 2 scores, but the effect of this was minimal and age does not appear to operate as a barrier to, or an effective predictor of, success for either component. The differences between the performances of female and male candidates, and those of different ethnic origins, were also found to be so minor as to prevent the identification of any discernible trends across Part 1 or 2.

Across each Part 1 test administration and OSCE circuit some fluctuation was observed in the level of challenge and difficulty of the assessment items. This may be partly explained by the degree of variability in cohort ability. However, the analysis shows that there were also a number of indicators of stability in the inputs and outputs of each component. For Part 1, there was a great deal of consistency in the reliability coefficients, suggesting that the examination produces outcomes for candidates that are typically dependable and generalisable to other Part 1 examinations (in that candidates, on the whole, would be expected to achieve the same result with another, parallel version of Part 1). The facility and discrimination values for Part 1 were also within an acceptable range when analysing performance across the different blueprint areas. This suggested that each blueprint area makes a positive contribution to the measurement of applied medical knowledge.

Part 2 is a more complex assessment environment, with candidates' outcomes being nested within circuits, stations and examiners. It was noted that across the OSCEs delivered in 2015-2016, half of all the circuit mean scores fell within a range of just 2%, suggesting some consistency in the difficulty and delivery of this examination over the different administrations. Some of this consistency will be attributable to examiner performance, and although it was shown that some examiners inevitably leaned towards the award of higher scores, or lower scores (bearing in mind the difficulty of the station/s to which they were assigned), no real outlier examiners were identified.

The analysis presented in this report is intended to contribute to the quality assurance cycle, by providing evidence of any annual trends that can help inform the various processes that the PLAB teams oversee to ensure the quality of the assessments and candidate outcomes. This report has highlighted where items, stations and examiners are performing as intended or where follow-up action may be required. As a next step, it would be beneficial to identify how this information can be applied in the contexts of item/station writing, examiner training, content sampling, and standard setting to maintain the quality of the assessments in the future.

APPENDIX 1: PART 1 ITEM PERFORMANCE: all items used in 2015-2016 (underlined denotes an item that produced a facility value of zero on at least one occasion).

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
1	3	.75	.38
2	3	.95	.22
4	2	.80	.37
5	2	.63	.36
<u>9</u>	<u>3</u>	<u>.37</u>	<u>.19</u>
10	2	.37	.16
12	1	.44	.18
16	3	.66	.38
17	1	.78	.33
19	1	.86	.29
20	3	.52	.19
27	1	.64	.43
28	1	.96	.27
29	1	.46	.34
31	2	.97	.34
34	2	.75	.45
37	1	.39	.23
38	1	.74	.26
40	1	.84	.31
43	2	.85	.36
44	1	.51	.17
46	2	.90	.42
49	3	.83	.35
50	1	.55	.23
53	2	.79	.30
56	2	.57	.40
59	3	.80	.23

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
61	1	.54	.28
62	2	.66	.51
65	2	.76	.13
69	3	.79	.16
71	1	.86	.37
72	2	.27	.20
76	1	.20	-.02
79	1	.84	.22
80	1	.62	.26
83	2	.69	.31
88	2	.56	.12
92	1	.82	.40
94	1	.97	.22
95	2	.66	.39
102	1	.31	.00
103	2	.44	-.24
105	3	.76	.17
106	2	.49	.21
109	2	.88	.30
112	1	.10	.01
113	2	.69	.30
115	3	.48	.40
123	3	.76	.30
124	1	.70	.34
126	3	.77	.36
130	2	.87	.39
132	1	.90	.27

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
133	1	.82	.29
135	2	.24	.13
142	3	.69	.27
144	1	.22	-.04
149	3	.78	.40
150	1	.80	.36
153	1	.79	.29
154	2	.70	.28
156	2	.65	.29
161	1	.89	.30
162	3	.91	.34
165	3	.68	.37
166	3	.93	.34
169	1	.61	-.05
173	2	.69	.12
175	1	.82	.39
181	2	.33	.08
183	3	.69	.05
190	1	.76	.38
192	1	.62	.34
200	2	.76	.43
209	2	.68	.29
216	1	.69	.13
217	1	.61	.34
222	3	.63	.38
223	1	.91	.26
227	2	.89	.28

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
233	2	.93	.32
234	1	.91	.26
236	1	.96	.19
237	2	.45	.14
242	1	.54	0.1
243	4	.20	.07
248	1	.86	.20
254	1	.42	.25
257	1	.63	.40
273	3	.61	.14
274	1	.69	.37
277	2	.95	.26
279	1	.33	.31
292	2	.73	.33
299	2	.49	.25
305	1	.89	.33
308	2	.81	.34
309	4	.88	.39
310	1	.81	.22
311	1	.86	.26
312	1	.56	.07
314	2	.78	.32
317	1	.53	.27
320	1	.77	.38
325	1	.96	.18
326	2	.84	.23
332	1	.75	.30

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
335	1	.75	.43
336	2	.40	.12
337	2	.72	.18
341	2	.55	.10
344	1	.39	.02
345	1	.80	.30
348	1	.56	.32
351	1	.40	.10
359	1	.50	.08
360	2	.66	.27
373	1	.75	.38
376	2	.90	.29
378	1	.82	.21
385	2	.83	.32
393	2	.75	.24
397	2	.60	.20
401	1	.83	.41
402	1	.60	.34
403	1	.41	.27
405	1	.77	.28
417	2	.71	.29
418	1	.78	.26
420	1	.94	.23
437	1	.77	.38
442	1	.91	.06
443	3	.87	.29
444	1	.50	.30

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
446	1	.88	.38
448	2	.87	.30
451	1	.88	.16
455	2	.51	.32
460	1	.86	.39
462	1	.86	.39
463	3	.48	.20
464	1	.49	.36
468	3	.62	.42
470	2	.54	.36
471	1	.87	.34
472	1	.68	.38
473	3	.28	.10
479	1	.82	.36
481	2	.85	.28
484	2	.34	.13
486	1	.85	.18
493	2	.54	.12
494	2	.64	.34
495	1	.73	.43
496	1	.46	.24
501	1	.84	.30
503	1	.71	.36
504	3	.55	.03
514	1	.88	.27
517	2	.64	.34
518	1	.88	.31

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
523	3	.58	.03
524	1	.80	.18
526	1	.96	.32
527	3	.65	.34
530	3	.70	.48
536	1	.71	.03
537	2	.72	.25
538	1	.53	.27
541	1	.57	.40
542	1	.97	.25
552	3	.66	.35
556	1	.71	.22
560	2	.53	.12
561	3	.83	.30
564	2	.66	.29
581	1	.91	.41
585	1	.89	.41
590	1	.71	.38
591	1	.85	.27
593	1	.84	.32
607	2	.82	.29
608	1	.87	.14
611	1	.50	.41
613	1	.90	.26
614	2	.86	.20
615	2	.43	.25
618	2	.72	.37

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
624	2	.70	.24
630	2	.69	.38
635	1	.28	-.14
636	2	.20	-.03
642	2	.57	.48
647	3	.78	.24
650	2	.76	.15
651	1	.67	.32
655	1	.66	-.02
659	3	.78	.32
661	1	.77	.13
665	2	.93	.30
669	1	.59	.12
671	1	.19	.00
675	1	.78	.29
680	1	.63	.24
686	3	.59	.17
688	2	.62	.31
690	1	.83	.36
694	1	.43	.16
696	1	.69	.49
697	1	.57	.30
698	2	.84	.30
706	1	.53	.25
711	1	.40	.11
719	1	.84	.35
721	1	.57	.25

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
722	2	.44	.25
723	2	.80	.25
727	2	.62	.27
728	1	.55	.25
731	2	.75	.35
734	1	.42	.11
735	1	.86	.45
740	2	.70	.33
754	1	.91	.26
763	2	.86	.36
775	1	.74	.46
783	1	.33	.17
788	1	.92	.23
791	2	.40	.13
800	1	.53	.34
803	2	.14	.11
808	2	.06	.02
822	3	.75	.28
831	2	.90	.30
836	3	.67	.27
840	1	.28	.15
842	1	.55	.07
844	2	.52	.23
847	1	.90	.16
849	3	.64	.41
850	1	.97	.27
857	2	.58	.24

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
859	1	.96	.21
863	1	.59	.47
866	1	.38	.25
870	1	.39	-.10
875	2	.38	.08
885	1	.44	.17
886	2	.78	.32
888	1	.66	.23
891	2	.74	.42
893	1	.80	.35
896	2	.97	.20
898	1	.66	.38
901	2	.69	.30
903	1	.57	.28
908	2	.70	.17
910	1	.81	.34
911	1	.88	.46
913	3	.54	.34
914	1	.65	.30
925	2	.83	.36
931	1	.59	.19
938	1	.54	.20
940	1	.87	.24
941	2	.87	.32
944	1	.82	.38
951	2	.84	.39
957	1	.86	.38

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
969	3	.59	.10
970	2	.82	.32
971	1	.68	.23
972	1	.65	.30
977	1	.53	.21
981	1	.78	.41
984	1	.79	.31
986	1	.74	.27
999	1	.85	.25
1002	1	.28	.03
1006	1	.89	.37
1010	1	.84	.33
1013	2	.77	.26
1015	2	.76	.37
1016	1	.79	.18
1018	3	.70	.38
1030	1	.81	.53
1039	1	.85	.09
1066	1	.77	.36
1068	1	.82	.45
1069	1	.44	.30
1071	1	.07	.05
1082	1	.91	.32
1086	1	.53	.34
1092	2	.91	.29
1094	3	.53	.27
1108	1	.74	.40

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
1115	1	.67	.36
1123	1	.79	.18
1128	1	.87	.34
1130	1	.41	.29
1155	1	.87	.34
1156	3	.66	.39
1160	2	.62	.14
1166	1	.75	.39
1183	1	.82	.31
1184	2	.74	.26
1195	3	.22	.19
1204	1	.65	.34
1216	1	.91	.27
1220	1	.72	.38
1221	1	.98	.28
1225	1	.18	-.01
1263	2	.87	.36
1265	2	.77	.33
1266	1	.88	.42
1269	2	.72	.37
1270	3	.82	.37
1275	2	.57	.25
1277	1	.81	.32
1278	1	.81	.32
1284	1	.71	.39
1293	1	.64	.24
1300	1	.67	.13

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
1303	1	.81	.37
1312	1	.66	.08
1313	1	.83	.45
1317	1	.73	.33
1320	1	.85	.36
1325	1	.37	.14
1332	1	.78	.19
1334	4	.78	.29
1337	2	.72	.43
1345	1	.71	.29
1349	1	.20	.03
1350	1	.23	.08
1353	1	.79	.35
1356	1	.90	.32
1368	1	.83	.36
1369	1	.72	.44
1371	1	.43	.34
1372	1	.75	.29
1375	1	.72	.38
1378	1	.52	.32
1379	1	.93	.24
1381	1	.83	.38
1383	1	.73	.36
1386	1	.29	.17
1387	1	.68	.07
1396	1	.87	.18
1399	1	.34	.15

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
1406	2	.73	.33
1411	1	.76	.32
1414	1	.81	.28
1415	2	.82	.30
1419	1	.81	.36
1422	1	.37	.02
1425	3	.90	.26
1428	1	.54	.21
1434	1	.50	.24
1436	1	.79	.33
1447	1	.30	.24
1449	1	.64	.33
1450	1	.22	.09
1453	3	.35	-.06
1459	1	.16	-.20
1460	1	.67	.25
1463	1	.73	.23
1468	1	.87	.38
1474	1	.67	.35
1475	2	.77	.35
1476	1	.59	.27
1477	2	.78	.36
1478	1	.73	.38
1483	1	.94	.22
1486	1	.15	-.03
1487	1	.72	.27
1495	2	.58	.25

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
1496	1	.79	.52
1506	1	.68	.22
1511	2	.85	.20
1523	1	.73	.40
1525	1	.85	.31
1529	1	.73	.33
1534	1	.70	.33
1542	2	.68	.39
1543	1	.41	.25
1549	1	.63	.33
1551	1	.77	.15
1556	1	.75	.26
1558	2	.52	.34
1559	2	.08	-.05
1560	3	.66	.45
1561	1	.25	.18
1564	1	.23	.11
1568	2	.62	.39
1577	2	.82	.26
1582	1	.95	.37
1586	4	.94	.27
1594	1	.78	.46
1599	1	.73	.48
1600	1	.84	.32
1603	1	.33	-.05
1609	1	.76	.39
1611	2	.63	.10

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
1615	1	.64	.12
1621	1	.65	.32
1630	1	.55	.20
1642	1	.57	.09
1643	2	.59	.26
1655	2	.89	.35
1663	1	.73	.40
1665	1	.84	.38
1667	3	.26	.14
1675	1	.74	.22
1678	1	.28	-.12
1682	1	.67	.10
1683	1	.64	.00
1688	1	.45	.32
1690	1	.42	.11
1697	3	.50	.14
1706	1	.42	.17
1716	1	.78	.25
1717	1	.89	.28
1719	3	.47	.15
1730	1	.79	.18
1735	1	.84	.42
1736	1	.80	.37
1737	2	.93	.21
1738	1	.80	.26
1739	1	.77	.36
1741	1	.71	.29

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
1749	1	.69	.46
1766	1	.68	.32
1774	2	.73	.32
1775	3	.64	.37
1776	2	.70	.45
1781	1	.54	.16
1783	1	.09	.12
1787	1	.53	.39
1791	2	.78	.38
1793	1	.69	.15
1797	1	.94	.23
1799	1	.89	.41
1807	2	.36	.22
1817	1	.32	-.14
1823	1	.92	.27
1825	1	.79	.31
1826	1	.45	.09
1833	1	.64	.24
1834	1	.63	.31
1835	1	.47	.33
1836	2	.32	.06
1837	1	.82	.31
1839	1	.92	.25
1842	1	.74	.37
1843	1	.71	.20
1844	2	.67	.17
1845	1	.85	.08

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
1846	2	.92	.28
1847	1	.86	.44
1858	1	.59	.12
1864	1	.87	.31
1867	1	.46	.16
1868	1	.38	.15
1872	1	.78	.24
1880	4	.49	.44
1881	1	.96	.22
1882	1	.85	.39
1887	2	.82	.26
1889	1	.97	.26
1891	1	.70	.33
1896	1	.46	.11
1907	1	.95	.27
1909	1	.87	.18
1910	1	.58	.43
1914	1	.80	.24
1922	1	.70	.38
1933	1	.78	.30
1937	1	.74	.31
1942	1	.80	.35
1943	1	.71	.39
1944	1	.78	.20
1947	1	.52	.22
1954	1	.48	.25
1955	3	.85	.30

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
1956	1	.85	.26
1957	1	.89	.34
1958	2	.92	.15
1962	1	.86	.30
1963	1	.70	.25
1964	1	.81	.25
1966	1	.75	.42
1972	3	.79	.45
1973	3	.59	.21
1976	2	.80	.25
1977	2	.59	.48
1983	1	.83	.40
1987	2	.91	.28
1988	1	.68	.34
1990	1	.32	.05
1994	2	.27	.27
1995	1	.63	.11
2003	1	.84	.27
2004	2	.75	.22
2005	1	.47	.23
2006	3	.71	.40
2008	2	.88	.20
2012	1	.83	.41
2014	1	.69	.42
2019	1	.81	.37
2024	1	.76	.11
2027	4	.93	.34

Item ID	N test uses	Fac	Discr
Ave	1.5	0.67	0.26
2028	3	.90	.39
2029	4	.15	.05
2031	2	.63	.38
2033	1	.82	.38
2034	1	.74	.31
2035	1	.86	.39
2038	1	.68	.50
2040	1	.51	.36
2042	1	.30	-.08
2046	1	.71	.37
2047	1	.63	.23
2049	1	.89	.30
2052	1	.51	.24
2053	1	.19	.19
2054	1	.86	.30
2059	1	.26	.18
2063	1	.35	.03
2064	1	.89	.25
2066	1	.83	.28
2069	1	.33	-.12
2070	1	.46	.16
2072	1	.69	.12
2073	1	.17	-.09
2075	1	.59	.02

APPENDIX 2: PART 2 STATION PERFORMANCE: all stations used in 2015-2016

The following table shows the descriptive scores for each station used in 2015-2016. Please note that raw score data are used for the minimum and maximum values, and all titles are presented as they are recorded in the original data download, including where the title is incomplete. The facility value is the measure of station difficulty, but can also be read as the average achieved percentage score for that station.

Station UID	Station title	Facility	Discrim	Min (raw)	Max (raw)	N uses	Station fails (%)	Station passes (%)
Average		.58	.43	0.64	3.84	237	29	71
842	Myocardial infarction history	.67	.37	1.9	3.8	48	0.0	100.0
852	Constipation	.52	.23	0.8	3.9	41	43.9	56.1
854	Abdominal pain	.62	.34	1.2	4.0	273	35.9	64.1
855	Eye examination	.54	.48	0.1	4.0	591	41.5	58.5
858	Explaining treatment	.58	.36	0.8	4.0	136	30.9	69.1
860	Management of post-operative pain	.61	.54	0.4	4.0	195	23.6	76.4
862	Giving information on a surgical procedure	.51	.32	1.0	3.0	41	34.1	65.9
865	Telephone: Surgical Emergency	.57	.41	0.6	4.0	271	31.4	68.6
867	Rectal bleeding	.54	.42	0.3	3.2	133	25.6	74.4
869	Dysphagia	.53	.38	1.0	3.7	187	26.7	73.3
873	Overdose	.54	.51	0.1	4.0	316	39.9	60.1
874	Local anaesthesia	.57	.46	0.6	3.9	141	9.9	90.1
877	Pain relief - cancer	.53	.33	0.6	4.0	176	31.8	68.2
883	Crying baby	.55	.49	0.5	4.0	540	24.6	75.4
885	Fractured femur in a seven month old infant	.59	.43	0.4	4.0	313	26.5	73.5
886	Convulsion in young child	.72	.31	0.7	4.0	266	8.3	91.7
890	Diabetic check of the legs	.62	.39	0.9	4.0	150	14.0	86.0
892	Weight loss	.63	.32	0.8	3.9	177	24.3	75.7
895	Knee examination	.68	.47	0.3	4.0	182	14.3	85.7
898	Infant screaming attacks	.60	.37	0.3	3.9	131	24.4	75.6

Station UID	Station title	Facility	Discrim	Min (raw)	Max (raw)	N uses	Station fails (%)	Station passes (%)
Average		.58	.43	0.64	3.84	237	29	71
899	Child diarrhoea (telephone)	.63	.27	0.9	3.8	90	18.9	81.1
905	Head injury	.58	.43	0.7	4.0	144	23.6	76.4
907	Trauma - secondary survey	.50	.47	0.4	3.9	179	50.3	49.7
909	Knee injury - examination	.83	.30	1.5	4.0	41	7.3	92.7
914	Neurological examination of legs	.50	.26	0.3	3.8	312	46.8	53.2
919	Neurological examination of legs	.44	.20	0.7	3.5	140	73.6	26.4
922	Severe headache	.65	.39	0.8	3.9	189	17.5	82.5
923	Headache - explanation	.56	.54	0.7	4.0	127	29.1	70.9
924	Cranial nerve examination	.52	.31	0.5	4.0	89	42.7	57.3
929	Giving a positive result for a test for gono	.61	.40	0.9	4.0	194	19.1	80.9
930	History taking - STIs	.57	.53	0.8	4.0	132	31.8	68.2
933	Pre-operative assessment	.59	.60	0.8	4.0	48	31.3	68.8
934	Missed miscarriage	.57	.27	0.4	4.0	386	31.1	68.9
938	Pre-eclampsia	.62	.42	0.7	4.0	174	19.5	80.5
946	Bi-manual examination	.66	.42	0.0	4.0	188	20.2	79.8
947	Sterilisation	.56	.51	0.6	3.8	176	23.3	76.7
948	Ovarian cyst	.56	.37	0.6	3.8	137	33.6	66.4
949	Anxiety	.53	.54	0.7	3.7	232	40.5	59.5
950	Alcohol history	.57	.56	1.3	3.8	48	25.0	75.0
951	Weight loss	.50	.50	0.7	3.9	231	48.5	51.5
952	Confusion in an elderly person	.54	.36	0.5	3.6	630	37.6	62.4
954	Antidepressant therapy	.59	.39	0.8	4.0	325	30.8	69.2
955	Insomnia	.50	.37	0.7	3.7	181	41.4	58.6
962	Respiratory - history	.60	.31	1.0	3.9	137	38.0	62.0
963	Haemoptysis	.58	.35	0.8	4.0	305	43.9	56.1

Station UID	Station title	Facility	Discrim	Min (raw)	Max (raw)	N uses	Station fails (%)	Station passes (%)
Average		.58	.43	0.64	3.84	237	29	71
964	Suturing	.57	.47	0.0	4.0	894	28.1	71.9
967	Post-operative collapse (Telephone)	.63	.55	1.2	3.8	220	19.1	80.9
968	I/V Cannula insertion	.57	.46	0.1	4.0	758	31.9	68.1
972	Urethral catheterisation	.55	.30	0.0	4.0	861	46.6	53.4
978	Painful frequent urination with fever	.54	.36	0.1	4.0	415	35.2	64.8
2614	Dark motions - history taking	.62	.33	1.2	4.0	180	15.0	85.0
2616	Urinary incontinence	.56	.52	0.7	3.7	215	33.0	67.0
2624	History/management of lower abdominal pain	.56	.43	0.7	4.0	95	36.8	63.2
2642	Upper GI endoscopy - providing information	.51	.49	1.2	2.7	45	22.2	77.8
2841	Testicular lump - explanation	.58	.47	0.1	4.0	408	30.9	69.1
3024	History of opiate dependency	.57	.58	0.2	3.5	45	40.0	60.0
3087	Discharge Instructions	.59	.49	0.7	3.4	139	15.1	84.9
3158	Prognosis of pelvic inflammatory disease	.64	.44	0.4	4.0	88	18.2	81.8
3159	Counselling in multiple sclerosis	.51	.35	0.1	3.4	135	21.5	78.5
3169	Giving advice to a patient with poorly contr	.58	.48	0.3	3.8	174	21.3	78.7
3177	Explanation to parents of child with multipl	.49	.41	0.2	4.0	465	50.5	49.5
3180	Anaemia	.61	.33	0.9	3.8	312	24.4	75.6
3184	History-taking - drug abuse	.47	.60	0.8	3.5	46	60.9	39.1
3196	Long-term back pain	.57	.33	0.0	4.0	323	26.0	74.0
3197	Abnormal blood result	.55	.43	0.1	4.0	91	39.6	60.4
3204	Fever	.61	.38	0.9	4.0	91	15.4	84.6
3205	Antepartum haemorrhage	.55	.54	1.0	3.6	134	21.6	78.4
3220	Dizziness	.56	.34	1.2	3.2	135	17.8	82.2
3222	Obesity	.66	.61	1.2	3.6	47	8.5	91.5

Station UID	Station title	Facility	Discrim	Min (raw)	Max (raw)	N uses	Station fails (%)	Station passes (%)
Average		.58	.43	0.64	3.84	237	29	71
3233	Examination of lympho-reticular system	.58	.53	0.2	4.0	323	25.7	74.3
4013	Management of essential hypertension	.47	.39	1.1	2.6	43	46.5	53.5
4070	Hypoglycaemic seizure in childhood	.58	.48	0.7	4.0	358	25.1	74.9
4091	Double vision	.58	.51	0.8	4.0	209	26.3	73.7
4092	Gout	.54	.37	0.8	3.9	364	30.2	69.8
4100	Patient with a problem	.66	.34	1.0	4.0	96	13.5	86.5
4110	Self-harm	.54	.56	1.0	3.8	125	50.4	49.6
4122	Tiredness	.53	.40	0.4	3.3	225	37.3	62.7
4126	Elderly abuse	.62	.37	1.1	4.0	317	26.5	73.5
4127	Collapse	.67	.40	1.1	4.0	315	11.4	88.6
4128	Acute scrotal pain	.63	.45	0.5	4.0	415	21.0	79.0
4130	Paediatric gastrointestinal disorder	.54	.55	0.2	4.0	184	42.4	57.6
4132	Paediatric head injury	.59	.36	0.8	3.9	225	23.6	76.4
4133	Preparation of intravenous insulin infusion	.71	.46	0.0	4.0	541	35.9	64.1
4134	Allergic reaction	.57	.57	1.0	3.2	50	16.0	84.0
4137	Paediatric injury	.62	.46	0.4	4.0	140	19.3	80.7
4141	Osteoporosis	.58	.56	0.9	4.0	138	32.6	67.4
4209	Calf pain	.58	.64	0.9	3.5	230	19.1	80.9
4212	Painful right eye	.56	.49	0.8	4.0	179	31.3	68.7
4215	Blood transfusion	.59	.49	1.0	4.0	135	23.7	76.3
4445	Sexual history	.55	.47	0.9	4.0	142	38.0	62.0
4510	Explaining a diagnosis	.52	.54	0.0	3.7	274	28.8	71.2
4523	Headache	.57	.33	0.2	4.0	513	31.2	68.8
4530	Fractured neck of femur	.51	.37	0.3	3.8	358	30.4	69.6
4532	Care of the elderly	.57	.32	0.5	4.0	265	34.7	65.3

Station UID	Station title	Facility	Discrim	Min (raw)	Max (raw)	N uses	Station fails (%)	Station passes (%)
Average		.58	.43	0.64	3.84	237	29	71
4533	Diabetic retinopathy	.51	.39	0.8	3.3	137	38.0	62.0
4535	Joint pains	.64	.28	0.5	4.0	180	27.8	72.2
4587	Needlestick injury	.54	.40	0.4	3.8	226	38.1	61.9
4596	Headache	.64	.34	1.1	4.0	147	12.2	87.8
4620	Needlestick injury in a child	.71	.30	1.8	4.0	78	6.4	93.6
4621	Routine surgery	.53	.45	0.5	4.0	106	43.4	56.6
4626	Delayed walking in a twenty-four month old m	.61	.48	0.3	4.0	144	28.5	71.5
4627	Constipation	.63	.41	1.3	4.0	47	4.3	95.7
4630	History taking from an infertile woman	.65	.34	0.7	4.0	373	22.5	77.5
4878	Chest pain in a young adult	.61	.58	0.6	4.0	226	31.4	68.6
4885	Management of earache in a child	.57	.60	0.4	3.6	43	27.9	72.1
4886	Peak flow measurement	.57	.46	0.4	4.0	237	42.6	57.4
4887	Smoking cessation	.59	.40	1.1	3.6	97	18.6	81.4
4897	Temporal arteritis	.53	.57	0.2	4.0	262	31.3	68.7
4902	Difficulty in swallowing	.61	.41	1.3	4.0	183	23.0	77.0
4908	Abdominal pain	.57	.38	1.0	4.0	140	25.0	75.0
4945	Visual field examination	.60	.34	0.7	4.0	309	24.6	75.4
4998	Pain relief - cancer	.59	.27	0.7	3.9	183	14.2	85.8
5016	Needlestick injury	.44	.45	0.7	3.0	48	66.7	33.3
5143	Child with bruises	.63	.59	0.8	4.0	135	31.1	68.9
5459	Coma	.61	.51	0.2	4.0	308	36.7	63.3
5460	Arterial blood gases	.51	.57	0.2	3.8	44	38.6	61.4
5462	Cervical smear	.65	.42	0.3	4.0	436	22.2	77.8
5490	Breast examination	.57	.43	0.2	4.0	659	30.8	69.2

Station UID	Station title	Facility	Discrim	Min (raw)	Max (raw)	N uses	Station fails (%)	Station passes (%)
Average		.58	.43	0.64	3.84	237	29	71
5495	Epilepsy	.32	.50	0.5	2.5	47	87.2	12.8
5496	STDs - sexual history	.55	.35	0.6	4.0	134	41.8	58.2
5498	Haematuria	.56	.43	0.5	3.7	190	35.3	64.7
5499	Back pain	.62	.49	0.8	3.8	141	16.3	83.7
5749	Examination of the ear	.58	.41	0.3	4.0	646	30.8	69.2
5751	Acute chest pain	.64	.60	1.5	3.6	44	13.6	86.4
5757	Prostate examination	.69	.27	0.5	4.0	222	23.0	77.0
5942	Inhalation/ingestion of foreign body by todd	.60	.51	0.3	4.0	235	15.7	84.3
5944	Examination for thyroid disease	.63	.44	0.1	4.0	724	21.8	78.2
5946	Obstetric examination	.61	.38	0.5	4.0	266	29.3	70.7
5982	Trauma - primary survey	.54	.51	0.2	3.7	376	53.7	46.3
5983	Depressed patient	.57	.45	0.0	4.0	225	32.4	67.6
5984	Psychiatric risk assessment	.50	.30	0.1	4.0	234	41.9	58.1
5985	Management of alcohol dependency	.49	.42	0.7	3.5	303	39.9	60.1
5986	Management of dysmenorrhoea	.53	.27	0.8	4.0	87	43.7	56.3
5987	Rectal bleeding in a child	.73	.45	1.7	4.0	136	8.8	91.2
5988	Volume spacer	.64	.45	0.8	4.0	861	22.1	77.9
6011	History of falls	.61	.46	1.2	4.0	140	19.3	80.7
6013	Child with scald	.57	.31	1.1	4.0	188	26.1	73.9
6073	Diagnosis and treatment of painful elbow	.58	.51	0.8	3.9	184	38.6	61.4
6127	Contraceptive advice	.61	.46	1.1	4.0	93	18.3	81.7
6128	Advice on risk reduction	.63	.39	0.4	4.0	177	23.7	76.3
6170	Initial examination	.62	.48	0.4	4.0	494	29.4	70.6
6171	Painful right leg	.56	.41	0.3	3.9	786	33.6	66.4
6172	Chest pain	.58	.46	0.4	4.0	583	23.8	76.2

Station UID	Station title	Facility	Discrim	Min (raw)	Max (raw)	N uses	Station fails (%)	Station passes (%)
Average		.58	.43	0.64	3.84	237	29	71
6173	Hysterectomy	.63	.41	0.0	4.0	495	29.1	70.9
6183	Venepuncture	.61	.42	0.2	4.0	679	39.0	61.0
6184	Colonic cancer	.50	.45	0.7	3.5	91	46.2	53.8
6185	Colonic cancer	.58	.45	1.1	4.0	98	20.4	79.6
6186	Acute shortness of breath	.59	.46	0.1	3.9	268	18.7	81.3
6257	Coeliac disease	.62	.2	0.6	3.7	96	3.1	96.9
6258	History taking - breathlessness	.53	.29	0.9	3.3	46	47.8	52.2
*6356	Management of convulsing child	.63	.42	0.8	4.0	368	20.4	79.6
6357	Breaking bad news	.59	.46	1.0	3.8	43	27.9	72.1
6359	Neck injury	.57	.37	1.0	4.0	358	22.1	77.9
6360	Discharge Instructions - Asthma	.58	.37	0.9	4.0	141	29.1	70.9
6361	Examination of abdomen	.60	.39	0.3	4.0	506	26.1	73.9
6395	Stroke	.60	.38	0.7	3.6	237	19.4	80.6
6410	Wheeze	.60	.65	1.0	3.3	48	14.6	85.4
6411	Examination of patient with hearing problems	.57	.50	0.1	4.0	282	34.0	66.0
6412	Fever and vomiting in a young child	.56	.32	0.8	3.6	91	31.9	68.1
6471	Patient with a complaint	.59	.35	1.1	4.0	189	23.3	76.7

*Also titled Management of febrile convulsion.

APPENDIX 3: PART 2 EXAMINER PERFORMANCE

The following table records, for each Part 2 examiner active in 2015-2016, the number of candidates examined; the mean score across all stations examined; the proportion of station passes and fails; and the distribution of examiner global judgements. The blue shading indicates cases where an examiner awarded scores, results or judgements that were lower, or less frequent, than one standard deviation of the mean for that variable, indicating a tendency towards the award of lower scores, or lower frequency of results or judgements. Red shading denotes that the examiner awarded scores, results or judgements that were more than one standard deviation above the mean, indicating a tendency to award higher than average scores, or greater frequency of results or judgements than average across all examiners.

There is inevitably some variability in the scores, results and judgements awarded by the examiners, and some of this will be attributable to the difficulty of the station they were assigned, the ability of the cohort they examined, as well as their own marking tendencies. Where an examiner is above or below the mean, by more than one standard deviation, across one or more of the measures in the table below, the opportunity to participate in an additional calibration exercise, and to compare their marking with other examiners, may be beneficial.

Examiner	N	Mean	SD	Min	Max	Station result		EOJ		
						Fail	Pass	B	F	P
Average	239	2.34	0.68	0.6	3.84	28.8	71.4	22.1	17.6	60.7
1	46	2.23	0.61	1.1	3.7	23.9	76.1	19.6	8.7	71.7
2	489	2.09	0.61	0.4	3.8	32.7	67.3	15.7	15.5	68.7
3	592	2.33	0.93	0.0	4.0	34.3	65.7	17.9	39.7	42.4
4	622	2.29	0.7	0.0	4.0	29.9	70.1	24.3	18.5	57.2
5	387	2.24	0.78	0.0	4.0	31.5	68.5	21.7	17.6	60.7
6	945	2.29	0.62	0.1	4.0	30.7	69.3	25.2	17.1	57.7
7	269	2.29	0.8	0.2	4.0	30.9	69.1	18.2	18.2	63.6
8	352	2.36	0.93	0.2	4.0	29.0	71.0	25.0	27.6	47.4
9	417	2.20	0.66	0.3	4.0	34.3	65.7	14.9	20.6	64.5
10	232	2.46	0.77	0.4	4.0	21.6	78.4	17.7	15.1	67.2
11	397	2.35	0.67	0.0	4.0	22.2	77.8	15.4	20.7	64.0
12	94	2.59	0.58	0.4	4.0	10.6	89.4	11.7	2.1	86.2
13	195	2.35	0.94	0.0	4.0	28.7	71.3	28.7	15.9	55.4
14	267	2.07	0.45	0.7	3.4	41.2	58.8	19.1	21.0	59.9
15	476	2.08	0.58	0.0	3.9	37.6	62.4	29.8	21.6	48.5
16	169	2.04	0.66	0.8	4.0	42.0	58.0	23.7	25.4	50.9
17	770	2.66	0.92	0.0	4.0	22.6	77.4	14.5	18.3	67.1
18	481	2.04	0.71	0.2	4.0	49.1	50.9	17.0	29.5	53.4
19	89	2.29	0.5	0.9	3.7	21.3	78.7	14.6	7.9	77.5
20	273	2.17	0.69	0.4	4.0	38.8	61.2	2.9	33.7	63.4
21	360	2.14	0.82	0.1	4.0	42.2	57.8	18.1	36.7	45.3
22	45	2.62	0.63	1.0	3.7	17.8	82.2	8.9	4.4	86.7
23	632	2.09	0.86	0.0	4.0	40.2	59.8	22.3	18.4	59.3
24	224	2.02	0.56	0.6	3.5	42.4	57.6	38.4	23.7	37.9
25	272	2.21	0.86	0.1	4.0	40.1	59.9	11.8	32.4	55.9
26	487	2.13	0.63	0.3	3.9	33.3	66.7	24.4	12.5	63.0
27	330	2.74	0.77	0.0	4.0	12.7	87.3	3.6	4.8	91.5
28	408	2.49	0.78	0.7	4.0	26.7	73.3	20.8	13.7	65.4

Examiner	N	Mean	SD	Min	Max	Station result		EOJ		
						Fail	Pass	B	F	P
Average	239	2.34	0.68	0.6	3.84	28.8	71.4	22.1	17.6	60.7
29	365	2.38	0.66	0.7	3.7	27.1	72.9	24.4	13.2	62.5
30	600	2.33	0.55	0.4	3.9	23.3	76.7	35.2	9.7	55.2
31	680	2.41	0.78	0.0	4.0	30.6	69.4	11.9	31.5	56.6
32	693	2.12	0.57	0.7	4.0	35.5	64.5	31.7	29.1	39.1
33	460	2.03	0.66	0.0	4.0	35.9	64.1	19.3	16.3	64.3
34	308	2.48	0.69	0.1	4.0	24.0	76.0	1.0	21.8	77.3
35	676	2.11	0.57	0.0	3.9	33.0	67.0	24.1	19.7	56.2
36	223	2.33	0.84	0.3	4.0	30.5	69.5	12.6	19.3	68.2
37	275	2.09	0.74	0.5	3.9	43.3	56.7	30.2	28.0	41.8
38	43	2.52	0.62	1.2	3.6	18.6	81.4	11.6	14.0	74.4
39	352	2.03	0.63	0.3	3.7	43.2	56.8	29.0	29.8	41.2
40	509	2.12	0.77	0.4	3.8	43.0	57.0	18.3	26.7	55.0
41	278	2.50	0.66	0.8	4.0	21.6	78.4	19.8	7.2	73.0
42	277	2.12	0.58	0.8	3.5	33.9	66.1	32.5	20.6	46.9
43	539	2.55	0.73	0.3	4.0	22.3	77.7	11.5	11.9	76.6
44	354	2.22	0.76	0.1	4.0	36.4	63.6	24.9	13.6	61.6
45	94	2.58	0.92	0.3	4.0	22.3	77.7	18.1	18.1	63.8
46	427	2.58	0.67	0.5	4.0	16.9	83.1	13.1	5.6	81.3
47	268	2.23	0.63	0.7	3.8	29.5	70.5	14.9	25.7	59.3
48	226	2.28	0.64	0.4	3.8	27.4	72.6	17.3	9.3	73.5
49	495	2.33	0.81	0.1	4.0	37.0	63.0	23.0	22.6	54.3
50	602	2.36	0.71	0.4	4.0	24.6	75.4	18.4	10.8	70.8
51	41	2.07	0.87	0.8	3.9	43.9	56.1	26.8	19.5	53.7
52	238	1.86	0.68	0.2	4.0	60.9	39.1	27.3	24.8	47.9
53	174	2.42	0.87	0.0	4.0	36.2	63.8	14.9	25.3	59.8
54	47	2.29	0.66	0.5	3.2	23.4	76.6	14.9	34.0	51.1
55	136	2.66	0.82	0.7	4.0	15.4	84.6	14.0	11.0	75.0
56	89	2.29	0.75	0.5	3.5	33.7	66.3	25.8	28.1	46.1
57	141	2.27	0.81	0.3	4.0	29.1	70.9	27.0	12.1	61.0
58	130	2.53	0.64	0.9	4.0	14.6	85.4	20.0	10.0	70.0
59	219	2.35	0.53	1.1	3.9	18.7	81.3	14.2	11.4	74.4
60	472	2.49	0.49	1.1	4.0	16.9	83.1	22.2	5.7	72.0
61	683	2.36	0.77	0.4	4.0	29.1	70.9	16.8	11.7	71.4
62	268	2.26	0.62	0.7	4.0	28.0	72.0	12.7	10.1	77.2
63	185	2.52	0.91	0.3	4.0	24.9	75.1	18.9	21.6	59.5
64	48	2.56	0.45	1.5	3.8	6.3	93.8	12.5	2.1	85.4
65	44	2.93	0.7	1.5	4.0	9.1	90.9	9.1	4.5	86.4
66	181	2.64	0.67	1.0	4.0	15.5	84.5	13.8	14.9	71.3
67	408	2.28	0.71	0.2	4.0	33.3	66.7	37.7	11.8	50.5
68	83	2.35	0.77	0.7	3.8	24.1	75.9	27.7	13.3	59.0
69	441	2.32	0.53	0.7	4.0	17.2	82.8	22.4	13.6	63.9

Examiner	N	Mean	SD	Min	Max	Station result		EOJ		
						Fail	Pass	B	F	P
Average	239	2.34	0.68	0.6	3.84	28.8	71.4	22.1	17.6	60.7
70	143	2.30	0.87	0.4	4.0	36.4	63.6	18.9	9.1	72.0
71	87	2.00	0.76	0.2	3.8	51.7	48.3	8.0	31.0	60.9
72	44	2.78	0.73	0.9	4.0	13.6	86.4	20.5	4.5	75.0
73	52	2.18	0.55	1.3	3.4	42.3	57.7	36.5	5.8	57.7
74	229	2.19	0.81	0.3	4.0	38.4	61.6	21.4	28.8	49.8
75	870	2.46	0.85	0.3	4.0	33.3	66.7	20.5	21.3	58.3
76	54	2.41	0.78	0.5	3.8	24.1	75.9	22.2	11.1	66.7
77	48	2.24	0.35	1.4	3.0	10.4	89.6	10.4	2.1	87.5
78	369	2.28	0.7	0.3	4.0	28.2	71.8	7.0	20.3	72.6
79	47	2.10	0.69	1.0	3.9	46.8	53.2	21.3	31.9	46.8
80	102	2.59	0.82	0.6	4.0	23.5	76.5	26.5	21.6	52.0
81	105	2.84	0.72	1.0	4.0	17.1	82.9	7.6	13.3	79.0
82	417	2.44	0.58	0.8	4.0	16.3	83.7	8.4	12.7	78.9
83	258	2.38	0.53	1.2	4.0	17.1	82.9	17.1	10.1	72.9
84	196	3.00	1.02	0.3	4.0	16.8	83.2	11.2	20.9	67.9
85	100	2.68	0.7	1.3	4.0	15.0	85.0	10.0	11.0	79.0
86	48	2.66	0.69	0.6	3.9	14.6	85.4	25.0	10.4	64.6
87	222	2.26	0.82	0.4	4.0	31.5	68.5	38.3	27.9	33.8
88	102	2.87	0.49	1.3	3.8	6.9	93.1	14.7	3.9	81.4
89	199	2.48	0.63	0.6	4.0	18.1	81.9	16.1	6.0	77.9
90	501	2.75	0.77	0.3	4.0	16.6	83.4	15.0	16.8	68.3
91	411	2.20	0.82	0.3	4.0	42.3	57.7	41.4	25.3	33.3
92	47	2.03	0.86	0.5	3.9	46.8	53.2	19.1	36.2	44.7
93	379	2.06	0.58	0.2	3.4	41.7	58.3	28.2	22.7	49.1
94	547	2.39	0.73	0.6	4.0	29.3	70.7	19.2	24.1	56.7
95	412	2.77	0.7	0.6	4.0	15.5	84.5	13.6	11.4	75.0
96	87	2.43	0.64	1.0	3.9	24.1	75.9	14.9	11.5	73.6
97	495	1.74	0.53	0.1	4.0	61.0	39.0	51.5	26.3	22.2
98	485	2.32	0.78	0.3	4.0	27.0	73.0	31.3	19.4	49.3
99	90	2.14	0.73	0.7	4.0	42.2	57.8	25.6	28.9	45.6
100	293	2.53	0.62	0.2	3.8	17.4	82.6	24.6	11.9	63.5
101	139	1.75	0.53	0.4	2.9	61.2	38.8	36.7	22.3	41.0
102	93	2.04	0.65	0.4	3.5	49.5	50.5	24.7	25.8	49.5
103	273	2.45	0.7	0.8	4.0	24.2	75.8	41.8	14.3	44.0
104	90	2.64	0.82	0.0	4.0	40.0	60.0	28.9	21.1	50.0
105	186	1.90	0.7	0.4	3.5	50.5	49.5	26.3	36.0	37.6
106	42	2.56	0.55	1.1	3.7	28.6	71.4	57.1	19.0	23.8
107	86	2.40	0.6	0.8	3.8	18.6	81.4	33.7	17.4	48.8
108	139	1.98	0.6	0.7	3.5	47.5	52.5	36.7	23.0	40.3
109	143	2.28	0.45	1.1	3.0	18.9	81.1	15.4	8.4	76.2
110	230	2.61	0.69	0.7	4.0	14.3	85.7	44.3	10.9	44.8

Examiner	N	Mean	SD	Min	Max	Station result		EOJ		
						Fail	Pass	B	F	P
Average	239	2.34	0.68	0.6	3.84	28.8	71.4	22.1	17.6	60.7
111	311	2.53	0.65	1.0	4.0	15.8	84.2	30.5	8.0	61.4
112	47	2.05	0.66	0.8	3.5	38.3	61.7	19.1	38.3	42.6
113	47	2.70	0.76	1.3	4.0	19.1	80.9	34.0	8.5	57.4
114	91	2.61	0.67	1.1	4.0	17.6	82.4	26.4	13.2	60.4
115	137	2.54	1.04	0.3	4.0	34.3	65.7	23.4	26.3	50.4
116	48	2.30	0.26	1.9	2.9	20.8	79.2	12.5	0.0	87.5
117	132	2.16	0.69	0.8	3.7	40.2	59.8	21.2	32.6	46.2
118	99	2.62	1.08	0.0	4.0	36.4	63.6	26.3	19.2	54.5
119	54	2.26	0.51	1.0	3.4	42.6	57.4	20.4	13.0	66.7
120	83	2.33	0.75	0.7	4.0	27.7	72.3	20.5	20.5	59.0
121	51	2.15	0.81	1.0	3.8	35.3	64.7	17.6	25.5	56.9
122	190	2.29	0.82	0.3	4.0	45.8	54.2	25.8	13.2	61.1
123	95	2.81	0.89	0.4	4.0	13.7	86.3	17.9	8.4	73.7
124	41	1.99	0.51	0.5	3.3	46.3	53.7	46.3	22.0	31.7
125	97	2.28	0.54	1.1	3.9	24.7	75.3	37.1	10.3	52.6
126	43	2.17	0.54	0.7	4.0	23.3	76.7	11.6	32.6	55.8
127	89	2.35	0.71	0.7	3.8	25.8	74.2	14.6	18.0	67.4
128	43	2.07	0.76	0.8	4.0	44.2	55.8	23.3	20.9	55.8
129	84	2.49	0.68	0.8	4.0	23.8	76.2	33.3	4.8	61.9
130	52	2.25	0.55	0.7	3.4	23.1	76.9	28.8	7.7	63.5
131	47	1.95	0.61	0.6	3.3	57.4	42.6	21.3	44.7	34.0
132	359	3.07	0.77	0.8	4.0	11.4	88.6	17.8	8.4	73.8
133	222	2.21	0.58	0.8	3.1	29.3	70.7	12.6	36.9	50.5
134	52	2.18	0.59	1.2	4.0	38.5	61.5	26.9	5.8	67.3
135	361	2.06	0.77	0.3	4.0	42.9	57.1	24.7	31.9	43.5
136	316	2.19	0.73	0.1	4.0	35.1	64.9	7.0	21.5	71.5
137	52	2.22	0.74	0.4	4.0	28.8	71.2	36.5	17.3	46.2
138	43	2.1	0.5	1.3	3.1	44.2	55.8	25.6	23.3	51.2
139	242	2.63	0.46	0.5	3.8	9.1	90.9	25.2	6.2	68.6
140	146	2.37	0.62	1.1	4.0	22.6	77.4	17.8	5.5	76.7
141	44	2.57	0.56	1.4	4.0	15.9	84.1	2.3	9.1	88.6
142	46	2.23	0.57	1.0	3.0	13.0	87.0	28.3	13.0	58.7
143	174	2.61	0.57	0.5	4.0	10.3	89.7	23.6	7.5	69.0
144	52	2.34	0.37	1.8	3.5		100.0	9.6	0.0	90.4
145	135	2.53	0.77	0.8	4.0	27.4	72.6	14.8	11.9	73.3
146	43	2.21	0.69	1.0	3.4	37.2	62.8	41.9	27.9	30.2
147	158	2.22	0.77	0.1	4.0	34.8	65.2	19.6	25.9	54.4
148	290	2.23	0.54	0.8	3.6	30.3	69.7	34.5	15.9	49.7
149	105	2.77	0.69	1.0	4.0	13.3	86.7	17.1	5.7	77.1
150	235	2.55	0.58	0.8	3.9	12.8	87.2	33.6	12.3	54.0
151	86	2.66	0.76	0.8	4.0	20.9	79.1	20.9	16.3	62.8

Examiner	N	Mean	SD	Min	Max	Station result		EOJ		
						Fail	Pass	B	F	P
Average	239	2.34	0.68	0.6	3.84	28.8	71.4	22.1	17.6	60.7
152	44	2.63	0.51	1.7	3.7	6.8	93.2	27.3	0.0	72.7
153	194	2.27	0.59	1.1	3.8	36.1	63.9	29.4	7.7	62.9
154	140	2.50	0.52	1.3	4.0	8.6	91.4	33.6	1.4	65.0
155	188	2.43	0.83	0.5	4.0	27.7	72.3	9.6	26.1	64.4
156	131	2.13	0.67	0.5	3.8	36.6	63.4	18.3	24.4	57.3
157	44	2.55	0.58	0.7	3.4	6.8	93.2	43.2	6.8	50.0
158	133	2.34	0.54	1.3	4.0	21.1	78.9	30.8	14.3	54.9
159	54	1.83	0.71	0.0	3.7	50.0	50.0	22.2	13.0	64.8