

A review of the dynamic setting of the stations-to-pass requirement for the CPSA:

PLAB Annual Report February 2026

Dr Matt Homer, Leeds Institute of Medical Education, University of Leeds

m.s.homer@leeds.ac.uk

Contents

List of tables	1
List of figures.....	2
Executive summary	3
Study overview	3
Key findings.....	3
Conclusions	3
Introduction	4
Methodology.....	4
The data samples	4
Setting the dynamic standard.....	5
Methods of analysis.....	6
Findings	7
The number of stations per exam.....	7
The unrounded dynamic standard.....	7
The rounded dynamic standard.....	10
Candidate-level pass/fail decisions.....	13
Comparison with earlier modelled PLAB2 data	16
Additional technical analyses.....	17
Discussion.....	18
References	19

List of tables

Table 1: Summary of station numbers per exam.....	7
Table 2: Summary statistics for dynamic standards	13
Table 3: Candidate overall pass/fail decisions under dynamic standards	14
Table 4: Candidate overall pass/fail decisions under fixed standards.....	14
Table 5: Candidate overall pass/fail decisions - dynamic versus fixed standards	15
Table 6: Candidates with different decisions under the two different rubrics	16
Table 7: Summary additional statistics for dynamic standards	18

List of figures

Figure 1: Example of calculation of dynamic standard	6
Figure 2: The unrounded dynamic standard for exams with 14 stations	8
Figure 3: The unrounded dynamic standard for exams with 15 stations	8
Figure 4: The unrounded dynamic standard for exams with 16 stations	9
Figure 5: The unrounded dynamic standard as a percentage of stations in the exam	10
Figure 6: The rounded dynamic standard for exams with 14 stations	11
Figure 7: The rounded dynamic standard for exams with 15 stations	11
Figure 8: The rounded dynamic standard for exams with 16 stations	12
Figure 9: Key data slide from Ottawa assessment conference presentation 2024 (Hankins et al., 2024)...	17

Executive summary

Study overview

In January 2024, the CPSA element of PLAB moved from having a fixed requirement in terms of stations needed to pass (e.g. 10 out of 16) to a dynamic standard that is designed to adjust with the difficulty of the set of stations in question (Figure 1).

Using CPSA data from January 2024 to November 2025 (page 4), this report evaluates how much the dynamic standard varies across CPSA administrations (page 7), and models the impact this change has had in terms of overall candidate pass/fail decisions in the exam (page 13).

Key findings

The typical standard for the full PLAB CPSA of 16 stations varies between 9 and 11 stations, with 9 and 10 stations predominating (Figure 8).

Compared to the previous fixed standard, the overall pass rate in the CPSA is 4.7 percentage points higher under the dynamic standard (Table 3 and Table 4). Overall, there are 4.7% of candidates who have different overall pass/fail decisions under the two systems, with the vast majority of these now passing under the dynamic standard when they would have failed under the fixed rubric (Table 5 and Table 6).

These results closely align with the modelling that was done prior to the introduction of the dynamic standard and suggest that the dynamic standard setting is working well and as intended (page 16).

More technical analysis suggests that the dynamic standard setting typically has good statistical metrics – for example, the R-squared between total domain scores and total stations passed shows a strong relationship upon which the method relies (Table 7).

Conclusions

Maintaining high quality and defensibility in OSCE-type examinations remains an on-going challenge. The analysis in this report suggests that the dynamic standard has increased the defensibility of PLAB CPSA outcomes in ways that were anticipated in advance of its introduction.

Monitoring of the dynamic standard should continue as part of the routine quality assurance processes in the CPSA, but there is no evidence in this report to suggest that further changes to the dynamic standard method are required, or that further additional focussed investigations of it are necessary.

Introduction

Many OSCE-type assessments require candidates to pass a minimum number of stations in addition to also achieving the overall pass mark. There are a range of arguments in favour of this additional requirement (Homer and Russell, 2021), but the most common justification given is to limit compensation across stations. In other words, to limit the extent to which candidates can do very well on a relatively small proportion of stations and then pass the exam without demonstrating a sufficiently broad set of knowledge and skills.

The inclusion of a minimum station requirement has natural appeal to a range of stakeholders, including patients, but historically the requirement has usually been set in each particular context to a fixed value (e.g. 10 out of 16 stations). This is clearly against best practice guidance in criterion-based assessment (Ben-David, 2000), where requirements for passing (e.g. cut-scores, minimum numbers of stations to pass) should adjust as the difficulty of the particular exam, made up of different sets of stations, varies from administration to administration. This adjustment ensures that the actual required standard (i.e. minimal competence) is maintained across test administrations. With a fixed standard, this is not the case.

A more defensible way of setting the minimum station standard, using post hoc analysis of examiner scores and grades from the OSCE, was proposed a few years ago (Homer, 2023). Following careful modelling and planning for implementation, this dynamic (i.e. not fixed) method was implemented in PLAB2 (now the CPSA element of PLAB) in January 2024. This report evaluates how the move to dynamic standard setting has operated over the period since then. It investigates what the actual standard looks like (i.e. how many stations are usually required to pass and how this varies from exam to exam) and the impact on overall candidate pass/fail decisions there has been compared to those under the previous fixed standard. It also compares these new data with what was expected given the modelling with earlier PLAB2 data that took place prior to the implementation of the new approach.

This report continues with a Methodology section that outlines how the dynamic method works, describes the data used in the report, and details the methods employed. The Findings give a comprehensive overview of what impact the move to the dynamic standard has had.

The report ends with a brief Discussion of what this work means for the ongoing development of the CPSA element of PLAB, and any implications for the wider assessment world within medical education.

Methodology

The data samples

Anonymised station-level candidate domain scores and grades were analysed from all PLAB CPSA exams over the period 05/01/24 to 04/11/25 inclusive. This comprises 549 exams in total with 32,234 candidates¹ and 8,697 individual station administrations (median number of candidates in an exam=62, and median stations per exam 16).

¹ Approximately 7,400 of these candidates appeared more than once in the data (i.e. candidates who resat the CPSA over the period in question) but, given the nature of the investigation (i.e. within exam), this does not impact on the analysis.

Setting the dynamic standard

Full details of how the dynamic standard can be set in Homer (2023) (method 'M4'), but in brief the method works as follows:

1. For the exam in question, a line of best fit is fitted to a graph of candidate scores (x , horizontal) versus numbers of stations passed (y , vertical).
2. This line is used at the overall cut-score for the exam² to estimate the typical number of stations passed by the borderline candidate – see Figure 1 as an exemplar – taken from a recent PLAB CPSA.
3. With appropriate rounding³, this gives the number of stations required to be passed by the minimally competent candidate.

² Calculated within each station using borderline regression (McKinley and Norcini, 2014), and then aggregated to the exam level to produce the overall cut-score for the exam

³ Currently PLAB CPSA uses ceiling rounding – i.e. always rounding up any fractional amount to the next whole number.

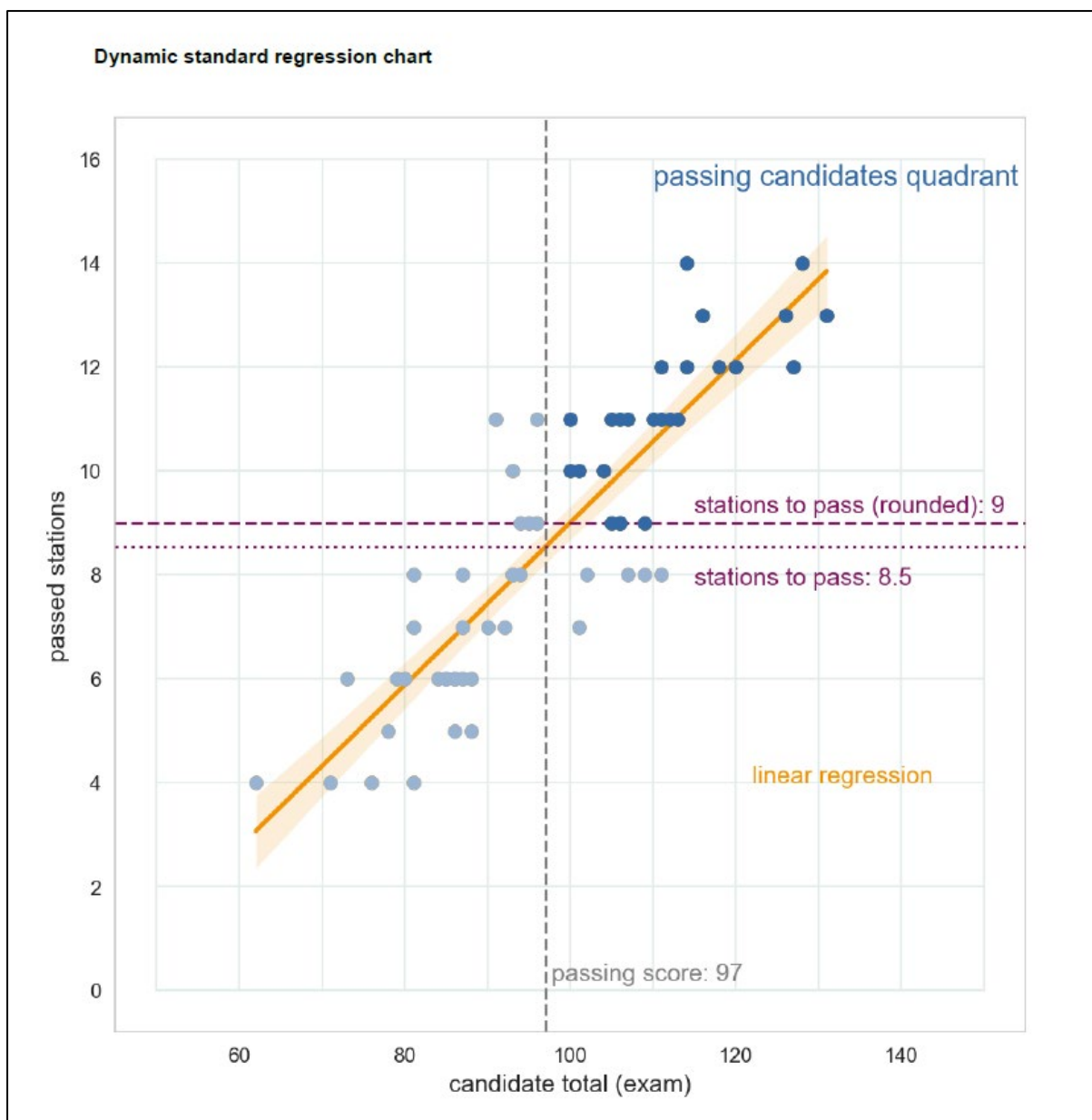


Figure 1: Example of calculation of dynamic standard

Methods of analysis

Using the sample of candidate scores and grades detailed above, the following calculations/analyses are carried out:

- within-station borderline regression cut-scores⁴
- cut-scores at the exam level (including the addition of 1 standard error of measurement calculated using the exam reliability and total score standard deviation)
- exam-level dynamic standards using the method described above (e.g. as per Figure 1)

⁴ With the borderline grade coded as 1, the current within-station borderline regression cut-score is set at 1.2 (i.e. just above borderline).

- individual candidate pass/fail decisions – based on passing the two separate requirements (the overall cut-score for the exam and the dynamic standard for stations)
- a comparison of individual candidate pass/fail decisions based on the dynamic standard and that under the previous rubric – which was 10 stations out of 16⁵

Findings

The number of stations per exam

On occasion, relatively poorly performing stations are suppressed (i.e. removed) from the CPSA after the exam is completed – based on the argument that the validity of the assessment and associated pass/fail decisions are improved without the inclusion of the station.

Table 1 shows a breakdown of the number of stations in each of the PLAB CPSA exams in the sample data. The vast majority (85%) consist of 16 stations. In what follows, it sometimes makes sense to present findings based on the number of stations in the exam.

Number of Stations	Frequency	Percent
14	4	0.7
15	79	14.4
16	466	84.9
Total	549	100

Table 1: Summary of station numbers per exam

The unrounded dynamic standard

The results here are presented in turn by the number of stations in the exam – 14, 15 or 16.

Figure 2 shows the distribution of unrounded dynamic standards for the very small proportion of exams consisting of 14 stations (i.e. each with two stations suppressed).

⁵ And 10 out of 15 stations if a single station is suppressed, and 9 out of 14 for the rare occasions where two stations are suppressed.

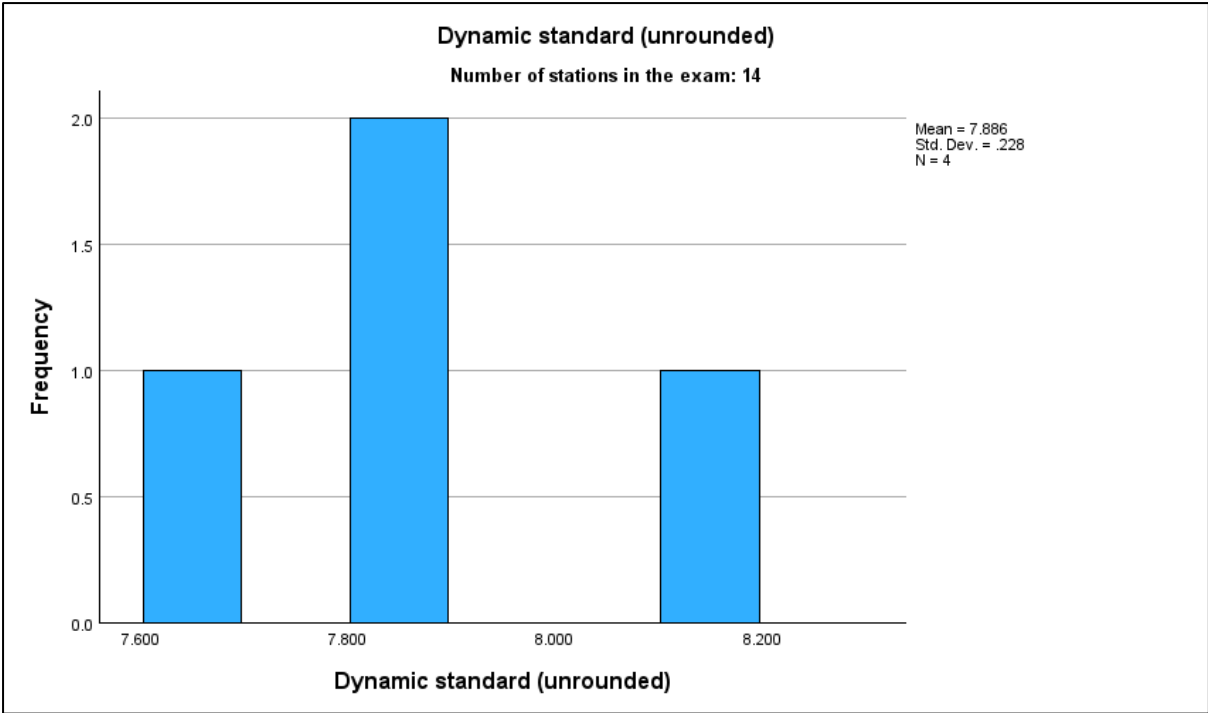


Figure 2: The unrounded dynamic standard for exams with 14 stations

Figure 3 shows the distribution of unrounded dynamic standards for the 79 exams with 15 stations (i.e. a single station suppressed) – the standard has a mean of 8.4 stations.

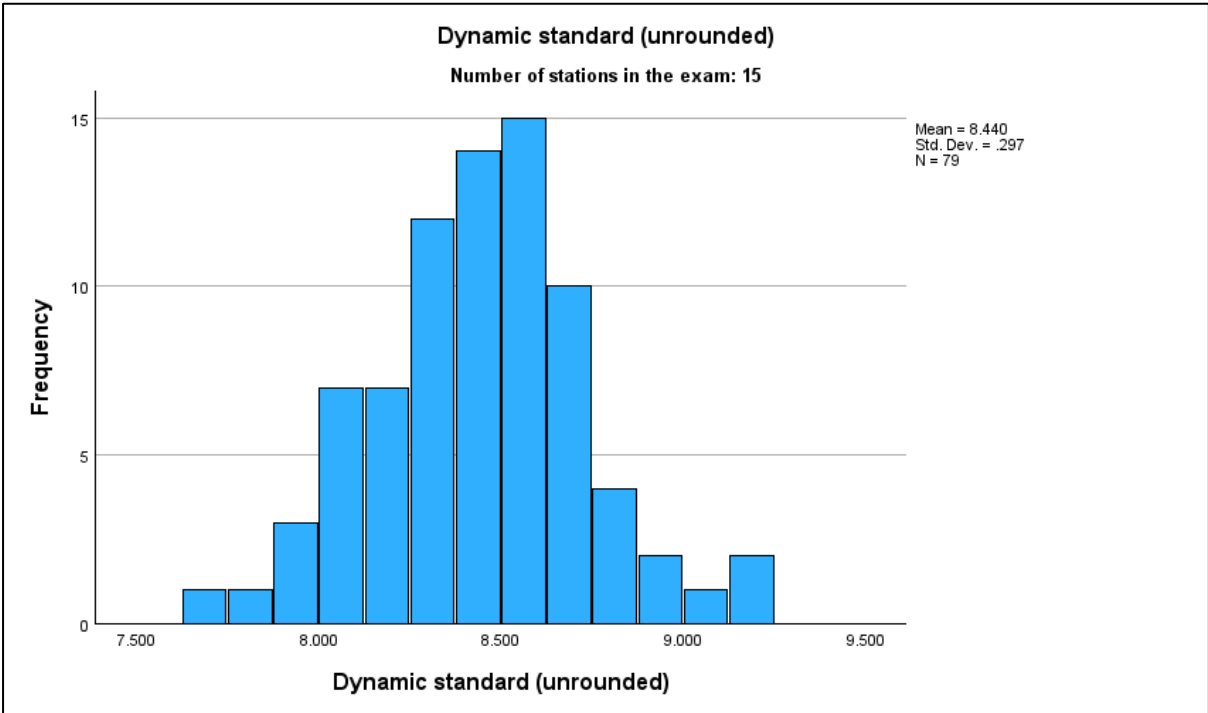


Figure 3: The unrounded dynamic standard for exams with 15 stations

Figure 4 shows the distribution of unrounded dynamic standards for the remaining 466 exams with a full complement of 16 stations – with a mean 9.0 stations.

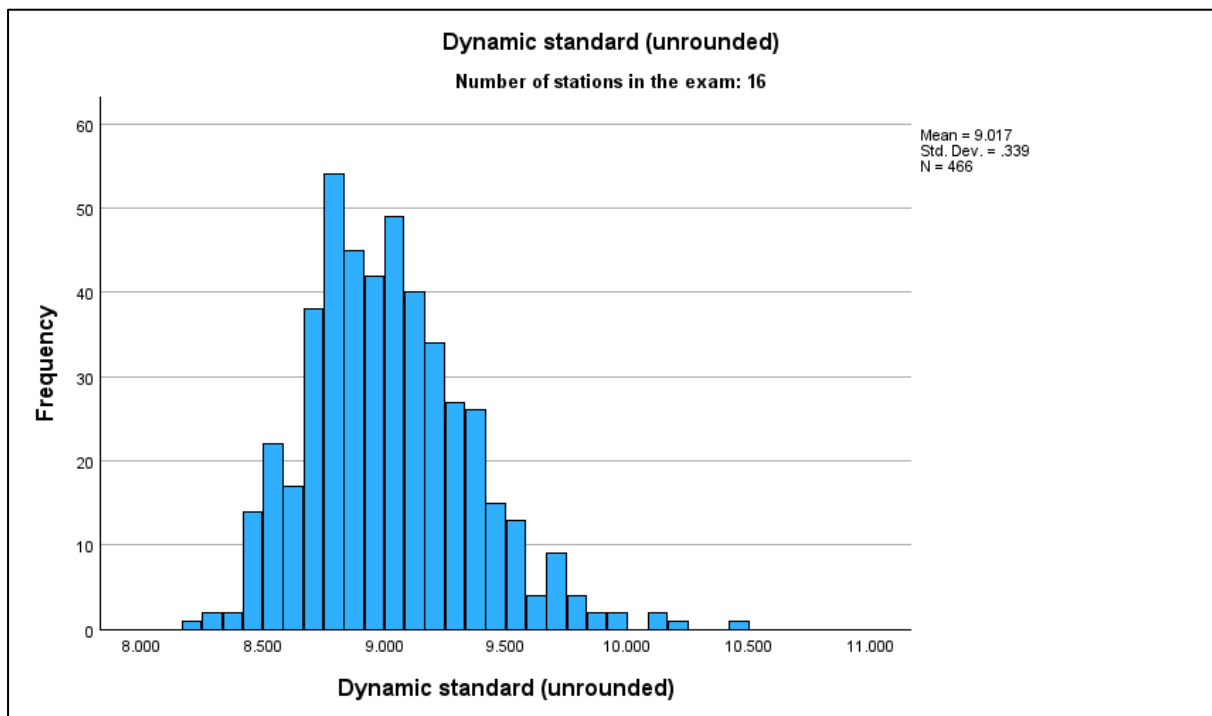


Figure 4: The unrounded dynamic standard for exams with 16 stations

Calculating the dynamic standard as a percentage of the stations in the exam means that the analyses above can be combined – see Figure 5 – with 56.3% of stations as the mean dynamic standard across the full dataset (minimum=51.3%, maximum 65.4%, median 56.2%).

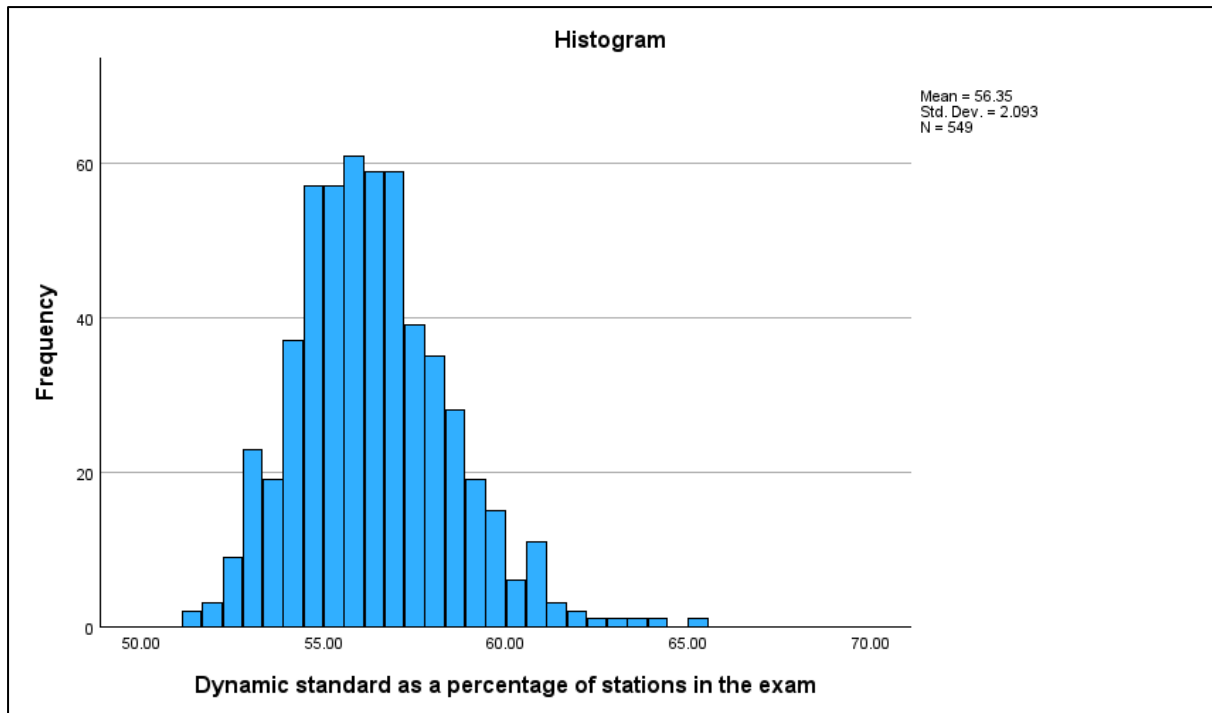


Figure 5: The unrounded dynamic standard as a percentage of stations in the exam

The rounded dynamic standard

As already mentioned, PLAB uses ceiling rounding to round up the decimal values summarised in the previous section. Again, this section is presented in turn by the number of stations in the exam – 14, 15 or 16.

The analysis begins with Figure 6 for 14 stations which shows that the dynamic standard is either 8 or 9 stations for 14-station PLAB CPSA exams – with 8 more common.

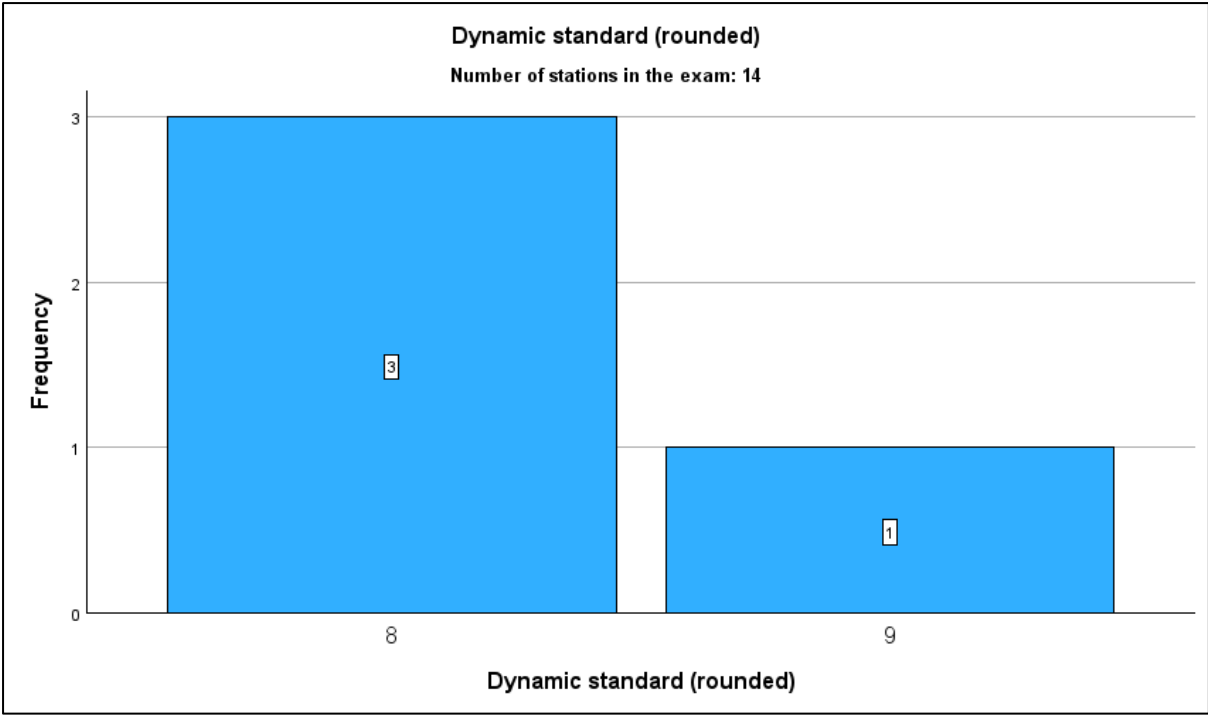


Figure 6: The rounded dynamic standard for exams with 14 stations

Moving on to 15 station exams, Figure 7 shows that 9 stations dominate as the standard (90.1% of exams), but there are small proportions with a dynamic standard of 8 or 10 stations (6.2% and 3.7% respectively).

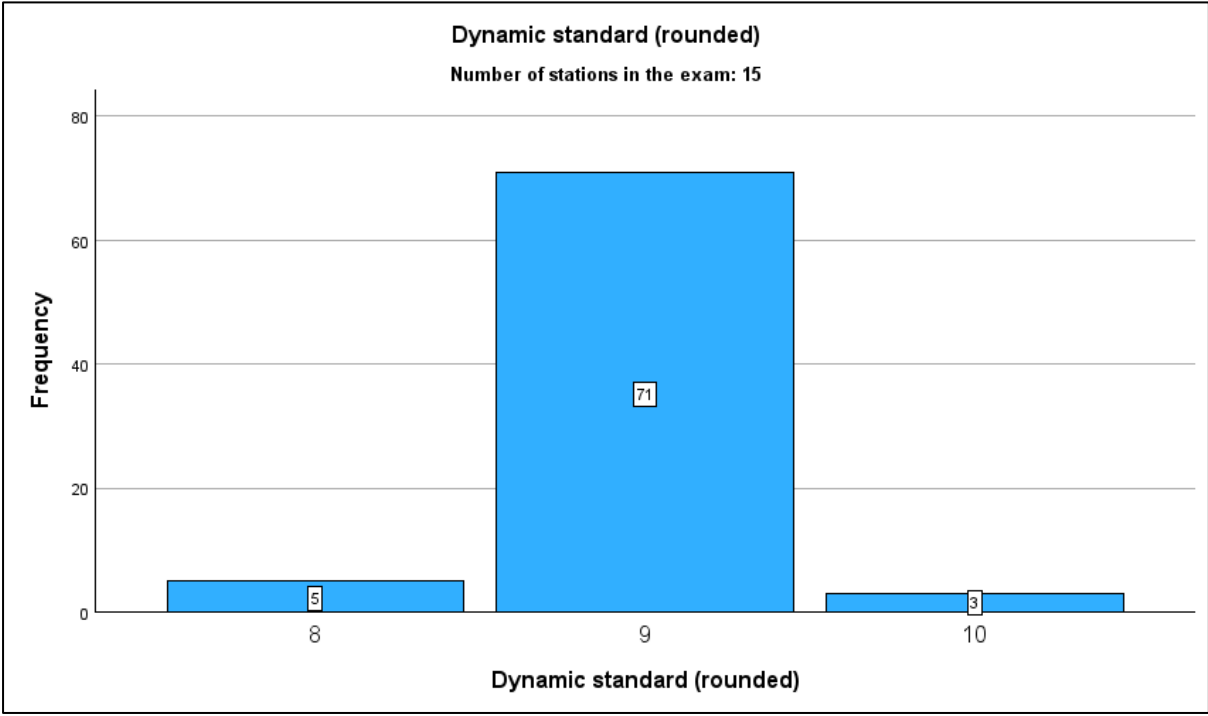


Figure 7: The rounded dynamic standard for exams with 15 stations

Finally, for the majority of CPSA exams with 16 stations, Figure 8 shows that 9 and 10 stations are the most common dynamic standards (51.3% and 48.1% of the exams) with, rarely, the standard being 11 stations (0.6% of exams).

This means that the dynamic standard is lower than the fixed standard would be (10 stations) in over half of recent 16-station PLAB CPSA exams.

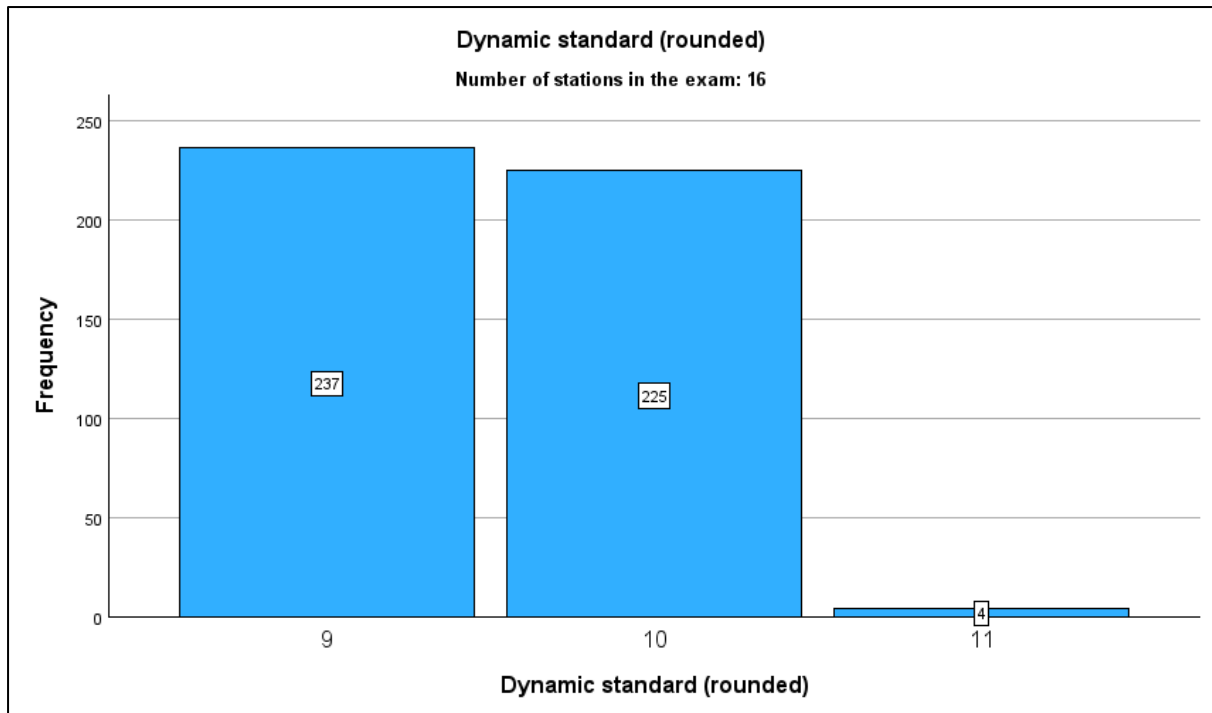


Figure 8: The rounded dynamic standard for exams with 16 stations

For completeness, Table 2 gives summary statistics for the dynamic standard across all PLAB CPSA data used in this report.

Number of stations in the exam	Statistic	Dynamic standard (unrounded)	Dynamic standard (rounded)	Dynamic standard as a percentage of stations in the exam
14	N	4	4	4
	Minimum	7.611	8	54.360
	Maximum	8.170	9	58.360
	Mean	7.886	8.250	56.326
	Std. Deviation	0.228	0.500	1.632
	Median	7.881	8	56.293
	Skewness	0.124	2	0.124

15	N	79	79	79
	Minimum	7.697	8	51.320
	Maximum	9.190	10	61.270
	Mean	8.440	8.970	56.267
	Std. Deviation	0.297	0.319	1.979
	Median	8.440	9	56.267
	Skewness	0.102	-0.564	0.102
16	N	466	466	466
	Minimum	8.212	9	51.320
	Maximum	10.457	11	65.360
	Mean	9.017	9.500	56.359
	Std. Deviation	0.339	0.517	2.119
	Median	8.987	9	56.171
	Skewness	0.643	0.187	0.643
Total	N	549	549	549
	Minimum	7.611	8	51.320
	Maximum	10.457	11	65.360
	Mean	8.926	9.42	56.345
	Std. Deviation	0.399	0.536	2.093
	Median	8.917	9	56.184
	Skewness	0.078	0.199	0.580

Table 2: Summary statistics for dynamic standards

Candidate-level pass/fail decisions

For the 32,234 candidates in the data, Table 3 summaries all pass/fail decisions under the current rubric for PLAB CPSA – including the dynamic standard setting for the stations-to-pass requirement. The overall pass rate is 63.7%, with 5.9% of candidates failing on the dynamic standard alone (shaded row) having achieved the overall cut-score for the exam.

Candidate pass/fail decision	Frequency	Percent
Fail both requirements	7,990	24.80
Fail only on cut-score requirement	1,793	5.60
Fail only on station requirement	1,916	5.90
Pass both requirements	20,535	63.70
Total	32,234	100

Table 3: Candidate overall pass/fail decisions under dynamic standards

Had the previous fixed rubric had remained in place (e.g. 10 out of 16 stations), the overall pass rate would have been 59.0% with 10.6% of candidates failing on the fixed standard – see Table 4 (shaded row). In other words, fewer candidates are failing under the dynamic standard – as expected given earlier analysis has shown that the dynamic standard is typically equal to or lower than the fixed standard would be (see, for example, Figure 8).

Another important finding is that the percentages of candidates failing based on only one of the two requirements is more balanced under the dynamic standard (Table 3) compared to the fixed one (Table 4).

Candidate pass/fail decision	Frequency	Percent
Fail both requirements	8,815	27.3
Fail only on cut-score requirement	968	3.0
Fail only on station requirement	3,431	10.6
Pass both requirements	19,020	59.0
Total	32,234	100

Table 4: Candidate overall pass/fail decisions under fixed standards

A candidate-level comparison can also be made between individual pass/fail decisions under the two systems (dynamic versus fixed) – see Table 5 which compares in a cross-tab individual candidate pass/fail decisions at the exam level under the two systems.

		Fixed standards				Total
		Fail both requirements	Fail only on cut-score requirement	Fail only on station requirement	Pass both requirements	
Dynamic standards	Fail both requirements	7,984	6	0	0	7,990
	Fail only on cut-score requirement	831	962	0	0	1,793
	Fail only on station requirement	0	0	1,909	7	1,916
	Pass both requirements	0	0	1,522	19,013	20,535
Total		8,815	968	3,431	19,020	32,234

Table 5: Candidate overall pass/fail decisions - dynamic versus fixed standards

The candidates on the leading diagonal (green shading) are those for which all individual pass/fail requirement decisions are exactly the same under either system (92.7% of candidates).

Those shaded pink have the same overall fail decision under the two systems but differ in terms of the stations-to-pass requirement (failed in one system but passed in the other) (2.6%).

Those shaded red are the key group of those for which overall pass/fail decisions are different (4.7%) – the move to dynamic standards has changed their overall pass/fail decision for the CPSA. For the pink/red off-diagonal decisions in Table 5, additional interpretation is given in Table 6 – discussed below.

As expected, there are no candidates who fail only on the dynamic standard requirement in one system but fail on cut-score in the other (shaded grey in Table 5).

Description	N	Percentage
Candidates who fail under both rubrics but would pass the fixed station requirement	6	0.02
Candidates who fail under both rubrics but pass the dynamic station requirement	831	2.58
Candidates who would have failed under fixed system but pass under the dynamic system	1,522	4.72
Candidates who fail under the dynamic system but would have passed under fixed rubric ⁶	7	0.02

Table 6: Candidates with different decisions under the two different rubrics

The vast majority of candidates with different overall pass/fail decisions (1522 out of 1529, 99.5%) pass under the dynamic standard but would have failed under the fixed standard, with only 0.5% of these having the opposite results (7 out of 1529).

Comparison with earlier modelled PLAB2 data

Numbers of stations to pass

In early 2024, three months of PLAB2 data from September 2023 to November 2023 (81 exams) was used to prepare a talk by Richard Hankins and colleagues at the Ottawa conference in Melbourne, Australia (Hankins et al., 2024).

The talk was about preparing for the move from the fixed to dynamic standard in PLAB2, with the key data slide reproduced in Figure 9. This shows summary statistics and the distribution of modelled dynamic standards for the 16 station PLAB2 exam. This data can be compared to Figure 4 and the relevant data in Table 2.

The mean dynamic standards are very close (9.02 in current data, 9.06 in modelled data), and the spread in the current data is a little less (0.34 and 0.39 respectively).

⁶ All of these 7 cases were 16 station exams.

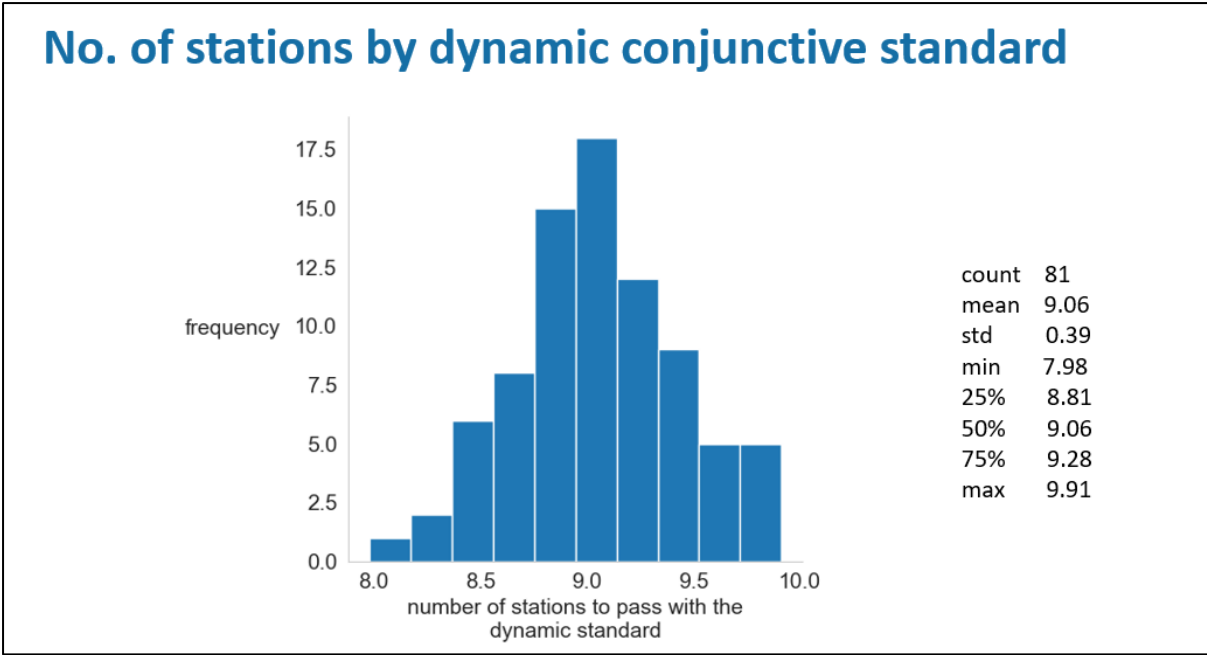


Figure 9: Key data slide from Ottawa assessment conference presentation 2024 (Hankins et al., 2024)

In other words, this comparison suggests that the actual dynamic standard has played out operationally in way very close to that expected based on the modelling carried out prior to the introduction of the dynamic standard.

Pass rates

Using ceiling rounding, the modelled data suggested that the pass rate from PLAB2 would increase by 3 percentage points (from 61% to 64%) (Hankins et al., 2024). The current analysis suggests that the move from fixed to dynamic standards has increased the pass rate from 59.0% (Table 4) to 63.7% (Table 5) - i.e. by 4.7 percentage points. This is a little higher than that predicted, but the comparison is not exact given that the degree of similarity in the profiles of candidates across the two time periods is unknown. In other words, there could be other (unknown) factors driving some of this difference rather than it be due solely to the introduction of the dynamic standard setting process.

Additional technical analyses

This section briefly documents the R-squared for each of the 549 PLAB CPSA exams, as well as the intercept and slope of the exam-level regression line (i.e. as per Figure 1) – see Table 7. These statistics might prove useful in future PLAB research, reports and ongoing quality control processes. Other institutions interested in implementing a similar approach to setting the minimum number of stations requirement will also be able to compare their metrics with these.

To allow all data to be aggregated across exams with differing numbers of stations, the intercept and slope are calculated based on percentage total scores and percentage total grades for each candidate.

Taking each column in Table 7 in turn, the R-squared figure for the calculation of the dynamic standard is typically around 0.78 which corresponds to a strong correlation between stations passed and total scores ($r=0.88$).

The typical intercept (mean) is -37.9% - whilst essentially not meaningful, this corresponds to the model-based percentage of stations required to pass for a overall cut-score of 0 percent.

The typical slope is a little easier to interpret – the mean value of 1.85 implies that a 1% increase in exam cut-score corresponds to an increase in the dynamic standard of 1.85 percentage points on average.

Statistic	R-squared	Intercept (%)	Slope (%)
N	549	549	549
Minimum	0.550	-71.960	1.150
Maximum	0.910	7.310	2.440
Mean	0.783	-37.916	1.854
Std. Deviation	0.057	10.868	0.192
Median	0.788	-38.250	1.848
Skewness	-0.678	0.236	-0.032

Table 7: Summary additional statistics for dynamic standards

Having completed the findings, the final section of the report is the Discussion.

Discussion

This work set out to investigate how the introduction of the dynamic standard setting approach in the PLAB CPSA has impacted on the minimum station requirement, and on individual candidate pass/fail decisions.

Typically, for the majority of CPSA exams with 16-stations, the dynamic standard does vary (between 9 and 11 stations, but with 9 and 10 stations predominating). This is the key finding in the sense that the method was designed to allow for adjustment of the minimum station requirement (Homer, 2023), and in practice this is exactly what happens. From a criterion-based assessment perspective (Ben-David, 2000), this makes the PLAB CPSA assessment outcomes more defensible than they were under a fixed standard.

The impact on candidate outcomes matches reasonably that expected when modelled in advance of the change of standard setting systems for the station requirement. As predicted, the pass rate is higher under the dynamic standard (by 4.7 percentage points).

With a fixed standard, the percentage of candidates failing only on the station requirement (i.e. having achieved the cut-score) is much higher compared to that under the dynamic standard (10.6% versus 5.9% respectively). One interpretation is that that this stations-to-pass standard might have been higher than ideal under the fixed method. Taking this argument further, under the dynamic approach there is closer alignment between the percentage of candidates failing on only one of the requirements (5.6% on cut-score only, and 5.9% on station requirement only). This improved

balance in the precise nature of pass/fail decisions is, arguably, additional evidence of the elegance and efficacy of the new approach to setting this standard.

All this evidence suggests that the dynamic standard setting approach is working well. The more technical analysis provides benchmarking for any future work and suggests that the dynamic method has good statistical metrics – e.g. when looking at the R-squared data.

As the PLAB CPSA develops, the dynamic standard (and other elements of the standard setting process) should remain under scrutiny. A similar analysis to that reported here could be repeated in time, but if results remain in line with those presented here this may not be necessary.

In terms of wider medical education assessment practices, it would be beneficial if other institutions where fixed station requirements are currently used would consider moving to employing the dynamic approach where possible. With appropriate modelling of data within their specific context, the evidence in this report suggests that a move to employing this approach would increase the defensibility of their own assessment outcomes.

References

- Ben-David, M.F. 2000. AMEE Guide No. 18: Standard setting in student assessment. *Medical Teacher*. **22**(2), pp.120–130.
- Hankins, R., Homer, M. and Caballero, J. 2024. Assessing a dynamic method of setting the conjunctive standard for a major licensing OSCE *In: Ottawa assessment conference, Melbourne.*
- Homer, M. 2023. Setting defensible minimum-stations-passed standards in OSCE-type assessments. *Medical Teacher*. **45**(10), pp.1163–1169.
- Homer, M. and Russell, J. 2021. Conjunctive standards in OSCEs: The why and the how of number of stations passed criteria. *Medical Teacher*. **43**(4), pp.448–455.
- McKinley, D.W. and Norcini, J.J. 2014. How to set standards on performance-based examinations: AMEE Guide No. 85. *Medical Teacher*. **36**(2), pp.97–110.