

The OSCE in a pre-registration osteopathy program: introduction and psychometric properties

Brett Vaughan^{1,2}
Prue Florentine¹

¹ College of Health & Biomedicine
Victoria University
Melbourne, Australia

² Institute for Sport, Exercise and Active Living
Victoria University
Melbourne, Australia

Corresponding author

Brett Vaughan
College of Health & Biomedicine
City Flinders Lane Campus
Victoria University
PO Box 14428
Melbourne VIC 8001
Australia

P. 61 3 9919 1210

F. 61 3 9919 1030

E. brett.vaughan@vu.edu.au

ABSTRACT

Objective

To investigate the psychometric properties of the objective structured clinical examination (OSCE) conducted in the final year of a pre-professional osteopathy program. A variety of metrics are used to determine the reliability of the examination. This is the first study to describe an OSCE in a non-USA osteopathy program.

Methods

Data from the OSCE conducted in 2011 was collated and analysed to establish the pass/fail rates, cost of the examination, internal consistency, and inter-examiner reliability. The examination was conducted over two days with students completing 5 stations on day 1 and 4 stations on day 2. Each station was of 15 minutes duration and there were 2 examiners per station.

Results

Forty-eight students and 31 examiners were involved in the examination. Twenty six students failed at least one station with six students failing three or more stations. Cronbach's alpha was greater than 0.80 for all stations indicating that each is internally consistent and over 50% of the variance in the students' total score for a station was due to the students themselves. Inter-examiner reliability was poor to fair however there

was good to excellent agreement for the global rating at each station. The total cost of conducting the examination was \$AUD12 933.20.

Conclusion

The results of the study suggest that the OSCE format is an appropriate method for assessing clinical competency in osteopathic education. The OSCE should be used in conjunction with other forms of assessment to develop an overall picture of the students' clinical competency. There are however some modifications that are required to improve the examination and these will be the subject of further studies.

Keywords

OSCE; reliability; competency assessment; educational measurement; osteopathy

The OSCE in a pre-registration osteopathy program: introduction and psychometric properties

ABSTRACT

Objective

To investigate the psychometric properties of the objective structured clinical examination (OSCE) conducted in the final year of a pre-professional osteopathy program. A variety of metrics are used to determine the reliability and validity of the examination.

Methods

Data from the OSCE conducted in 2011 was collated and analysed to establish the pass/fail rates, cost of the examination, internal consistency, and variance components. The examination was conducted over two days with students completing 5 stations on day 1 and 4 stations on day 2. Each station was of 15 minutes duration and there were 2 examiners per station.

Results

Forty-eight students and 31 examiners were involved in the examination. Twenty-six students failed at least one station with six students failing three or more stations. Cronbach's alpha was greater than 0.80 for all stations indicating that each is internally consistent and over 50% of the variance in the students' total score for a station was due to the students themselves. The total cost of conducting the examination was \$AUD12 933.20.

Conclusion

The results of the study suggest that the OSCE format is an appropriate method for assessing clinical competency in osteopathic education. The OSCE should be used in conjunction with other forms of assessment to develop an overall picture of the students' clinical competency. Some modifications to the format of the assessment are required to improve the examination and these will be the subject of further studies.

Keywords

OSCE; reliability; competency assessment; educational measurement; osteopathy

INTRODUCTION

The Objective Structured Clinical Examination (OSCE) is widely used in medicine and other health professions as a competency examination, and particularly assesses the *shows how* level of Miller's triangle.¹ The OSCE, in its basic form, is a multi-station examination where candidates demonstrate particular skills or knowledge at a station of fixed duration and marked by an examiner or examiners using set criteria. Improved reliability, particularly when compared to other forms of competency examination (i.e. long-case assessment), is one of the major issues driving the use of the OSCE in health professional assessment.

A number of authors have articulated the advantages of the OSCE when compared to other forms of assessment, and some of those factors relevant to an OSCE in osteopathic education include: high reliability and improved content validity;^{2, 3} increased number of skills being assessed;⁴ increased consistency in examinations between students;⁵ good inter-examiner reliability;² efficient method of examination;^{6, 7} cost effectiveness;⁸ and identification of areas of curriculum weakness.^{5, 9} Some of the disadvantages when using an OSCE in osteopathic education are similar to those in other health professions and include a reductionist approach;¹⁰ each station only assesses a small part of the clinical consultation;¹¹ limited ability to assess interpersonal skills;¹² stressful for students;^{2, 13} less reliable assessment of problem solving;^{14, 15} and limited assessment of clinical strategies due to no variability in patient presentation.¹⁶

From the original description of the OSCE by Harden¹⁷ numerous variations of the format of an OSCE have been published. These variations include the number of stations in the examination, the duration of each station, the number of examiners per station and the skills

and/or knowledge being assessed. McCrorie & Boursicot¹⁸ reported that within pre-registration medical programs in the UK alone, the format employed varied, the duration of an OSCE station varied between 4 to 45 minutes and the number stations employed in the exam between 6 to 48 stations. A systematic review by Walsh et al.¹⁹ identified nine studies examining an OSCE, with a range of between 8 to 42 stations and station duration of between 4 to 15 minutes. Each of these variations has an impact on the overall reliability of the examination and also makes comparisons between different OSCEs difficult. Overall reliability for the OSCEs identified in the review by Walsh et al.¹⁹ ranged from 0.40 to 0.91, with no pattern emerging as to whether more/less stations or increased/decreased station duration improved overall reliability.

Reliability of the assessment is impacted upon by the method chosen to 'mark' students. Assessments are scored on either an analytical metric, holistic scoring metric or a combination of both. Analytical metrics are typically checklists where the items on the checklist address the skills and processes undertaken during the assessment. Holistic metrics rate the assessment as a whole taking into account each element of the assessment.

Traditionally, an OSCE has been marked using analytical metrics (such as a checklist) where the examiner ticks off a list of elements that have been completed by the student. Whilst the checklist approach contributes to the high reliability of the examination, various authors have argued that this approach does not capture elements such as empathy^{20, 21} and undermines the expert clinical judgement by the examiner.¹³ This approach does provide the opportunity for structured feedback to the student post-examination.

Holistic metrics are used where complex or multidimensional constructs are being assessed. Some examples of these constructs include teamwork and communication. Quirk et al.²² concluded that global ratings are more reliable than checklists, and although it has previously been reported that global rating scales have poor reliability,²³ this may be related to lack of examiner training rather than the rating scale itself.

The cost of running an OSCE has been investigated by Walsh et al.¹⁹ These authors demonstrated a range of between USD\$6.90 to CA\$870 per student. Again, comparisons between OSCE's in relation to cost are difficult given the variations in format and number of students undertaking the examination, as well as whether standardised patients were used and the number of examiners who took part.

The current paper is not designed to provide an in-depth review of the OSCE as an assessment and readers are directed to Rushforth² or Walsh et al.¹⁹ for further information. The OSCE is used in medicine,¹⁸ dentistry,²⁴ nursing,^{25,26} physiotherapy,²⁷ chiropractic, and radiation therapy.²⁸ Whilst osteopathic teaching institutions report the use of such an assessment,²⁹ the investigation of the psychometric properties in non-US osteopathic education programs has not been documented.

The literature on competency assessment in osteopathy, both pre-registration and fitness-to-practice, is limited to commentaries²⁹⁻³² and the COMLEX examination in the USA for US-trained Doctor of Osteopathy graduates.³³⁻³⁶ No studies have been published on the use of an OSCE in a pre-registration osteopathy program. Rushforth² stresses that in evaluating an OSCE the trustworthiness of the assessment needs to be examined, that is, is the examination a reliable and valid assessment of competence? The aim of the current paper is to analyse the

metrics of an OSCE when used in the final year of the osteopathy program at Victoria University (VU).

METHOD

The OSCE has been used as part of the assessment of final year students at VU since 2005. Over this time, the process, content and station structure has evolved to that analysed as part of the present study. The study was approved by the Victoria University Faculty of Health, Engineering & Science Human Research Ethics Committee.

Examination format

Students were assessed at nine stations (Table 1) with the examination split into two days; five stations (plus two reading stations) on day one and four stations (plus 2 reading stations) on day 2. Each day was the split into two 'panels' where students would undertake the same station concurrently in order to improve the efficiency of the examination. Each station was of 15 minutes duration with 2 examiners per station. The total examination duration per student was two hours and fifteen minutes over the two days. A total of 8 clinical cases were developed and students undertook one case on day one and one case on day two, with each case containing clinical information relevant to each of the stations.

INSERT Table 1 here

Assessment format

Each criterion was graded using the scale in Figure 1 and a global pass/fail assessment for the station formed the grade awarded to the student. The total score for each student was not used to determine the global assessment. For the purposes of analysing the data, total marks were

collated for each student at each station and correlated with the global assessment to ascertain whether a pattern existed within the data.

INSERT Figure 1 here

Examiners

Examiners were recruited from the academic, casual and clinical teaching staff at VU. All examiners were provided with an examiners guide two weeks prior to the examination however no formal face-to-face examiner training was undertaken. The examiners guide contained information about the format and structure of the examination as well as a rubric about the performance level for each criteria. Examiners were briefed about the cases and the station content prior to the examination.

All examiners were paid AUD\$35.37 per hour with each examiner assessing for approximately 10 hours per day. Although tenured academic staff were required to examine as part of their normal duties, the cost associated with their involvement was calculated at the same rate as non-tenured academic staff.

Statistical analysis

The emphasis for the statistical analysis is the metrics of each individual station, as suggested by Fuller et al.³⁷ and follows the guide published by Pell et al.³⁸ Data from each mark sheet was entered into Microsoft Excel then transposed into SPSS for Mac. Global assessments

(Satisfactory or Unsatisfactory) were transformed into numerical values for the purposes of analysis.

Pass and fail rates were calculated. The total student score per examiner per station was calculated and correlated with the global assessment using a Pearson's r statistic. Pell et al.³⁸ suggest that the R^2 statistic is useful for indicating the relationship between the total score and the global rating. Subsequently, this statistic was also calculated. In addition, internal consistency was calculated for each station using Cronbach's alpha.

Variance components were calculated for each station with the student, examiner and case number treated as facets. A one-way ANOVA with VARCOMP was used to calculate the variance components. Alpha was set at $p < 0.05$. A generalisability analysis, although commonplace in the discussion of OSCE psychometrics, was not conducted as students did not have to pass a certain number of stations in order to pass the exam, therefore determining the overall reliability of the examination was not required.

The 'per student' cost was calculated and included examiner payments (includes both tenured and casual staff) and catering for the examiners on the day.

RESULTS

Participants

Fifty-one students were required to sit the 2011 OSCE at VU. The results of forty-eight final year osteopathy students were analysed. Two students results were not analysed; one student was unable to attend either day of the examination and one student was not required to complete all stations having completed a number of stations in the previous year.

Thirty-six examiners, including qualified osteopaths and medical practitioners, participated over the two days of the examination, with five of these examiners participating on both days. Thirty-two of the examiners had assessed a station in the OSCE in the previous years with the remaining four examiners not previously having been involved in an OSCE.

Pass & fail rates

Nine (19%) students failed one station, eleven (23%) failed two stations and six (12.5%) failed three or more stations. The average mark per station and the number of students who failed that station is presented in Table 2.

INSERT Table 2 here

The average total score awarded by the Lead Examiner and the Secondary Examiner were very similar for each station. Of note is the number of 'fails' in the Clinical Diagnosis and

Management station where slightly under one-third of students failed this station, with more students receiving a 'fail' grade in Panel 2 compared to Panel 1.

The correlation between the total score per student per examiner and the global assessment is presented at Table 3. Moderate correlations were achieved across all stations and only the Secondary Examiners in the Clinical Diagnosis and Management station achieved an R^2 value of greater than 0.50 suggesting a greater relationship between the criteria score and global rating awarded to the student.

INSERT Table 3 here

Inter-item reliability

The results of the internal consistency calculations for each of the nine stations are presented in Table 4. All stations are above the accepted alpha of 0.80 indicating that each station is internally consistent. When examining the 'if item deleted' statistics, two criteria when removed from the calculations, would result in a Cronbach's alpha that is higher than if the item were retained.

INSERT Table 4 here

Variance Components

The variance components for each station are presented in Table 5. Over 50% of the variance in the students score for a particular station is accounted for by the student themselves and in

five out of the nine stations, this variance was two-thirds or greater. When examining the Case History station, 27% of the variance was accounted for by the clinical history presented to the student. This is not an entirely unexpected result given that the student is expected to interpret the case history and develop a list of differential diagnoses based on the information presented.

INSERT Table 5 here

Cost

The total examiner payment for the two days was \$AUD12 733.20 and catering was AUD\$200 per day. The per-student cost was calculated to be AUD\$273.60 for the two days of the exam and this equates to AUD\$121.60 per hour of examination undertaken by the student.

DISCUSSION

The current paper, for the first time, describes the reliability and validity of an OSCE in a pre-professional osteopathy program. The format utilised for the examination described in this paper could be referred to as a 'long case' format, in that, the student followed one clinical case through on each day of the examination. The student was effectively assessed on two cases; one case on Day 1 and one case on Day 2. Each station on the OSCE was treated singularly for the purposes of the assessment, that is, a student who failed one station would only have to repeat that station, and not the entire examination.

Pass and fail rates for the examination were not compared to previous years' data given that the cases used in the 2011 OSCE were different to those used previously. The Clinical Diagnosis and Management station accounted for the most number of fails per station across the entire examination, with thirty percent of students attempting this station failing. Further analysis indicated that one panel failed more students than the other and the reasons for this are unclear. It is also important to look at this high-fail rate (in comparison to other stations). By examining the variance components it can be seen that the examiners accounted for 17% of the score variance on the station in addition to the 13% contributed by the student/examiner interaction variance. Given the data presented, a review of the station is warranted and modifications around the criteria for this station have been made for future OSCEs in an effort to standardise the station and limit the examiners to assessment of the performance of the clinical examination.

The Cronbach's alpha values for each of the stations was excellent. Only two criteria (Pharmacology criterion 4 and Rehabilitation criterion 1) when removed from the analysis

produced Cronbach's alpha values greater than the overall Cronbach's alpha for that station. Typically, this result suggests that the criterion may be measuring a different construct to the other criteria for that station. Of note is that these two criteria are addressing the medicolegal implications of both advice related to medicines use and advice related to the provision of rehabilitation exercises. The other criteria on these stations relate directly to the information contained in the case provided to the student and the two medicolegal criteria are generic and the response to these is not case specific. Whilst these criterion are particularly important from the point of view of assessing the students' knowledge of the medicolegal issues related to osteopathic practice, it may be that they are not appropriate for an OSCE given that it appears to be assessing a different construct, and subsequently should be assessed elsewhere. Further work is required to confirm that this is the case. The other stations produced more than acceptable internal consistency values.

Based on the findings of the present study as well as feedback received from both students and examiners who participated in this OSCE, a number of changes have been made to the structure of the assessment. A major improvement has been the blueprinting of the OSCE against the Victoria University Graduate Capabilities for Osteopathic Practice. This blueprinting has resulted in the addition of another station designed to assess the application of osteopathic manual techniques (in addition to the HVLA station in the current OSCE) as this was not being adequately assessed using other assessment methods. Another significant change to the criteria is the modification of the global assessment from a Satisfactory/Unsatisfactory (S or U) grade to a 4,3,2,1 scale reflecting that used to mark each of the criteria for that station. The interpretation of this scale is the same as that in Figure 1. This change has been made in order to differentiate between levels of performance on a station. Such a scale may produce substantial increases in the R^2 value over 0.50 indicating

that a high total station score is reflected in a high global score. It is likely that R^2 values over 0.50 are difficult to achieve when using a two-point scale (i.e. Satisfactory/Unsatisfactory) therefore the change to a 1-4 scale will also assist in providing feedback to the students about areas of their performance in which they can improve or conversely, areas in which their performance was above the standard expected.

The criteria for the Clinical Diagnosis & Management station have been modified to emphasise the assessment of a single systems examination (i.e. cardiovascular, respiratory) and to remove aspects of each criteria where the 'subjective' assessment of the examiner was relied upon too heavily. An osteopathic treatment station has also been added, taking the total number of stations for the 2012 OSCE to ten. Further, the Patient Management station has been retitled Patient Communication in order to emphasise and assess (in a simulated environment) the communication element of the osteopathic consultation. The Aftercare and Rehabilitation station criteria have also been modified slightly to remove any real or perceived overlap between the criteria at these stations.

The cost of the examination will increase in 2012 due to the addition of another station but will still be within the range of the identified by Walsh et al.¹⁹ The feasibility of an OSCE is an important consideration when deciding whether to use such an assessment and the cost data presented here may assist others in that decision. Whilst the cost of the assessment is quite high and resource intensive compared to other forms of performance-based assessments (i.e. mini-CEX) its value is in the reliability and fairness of the examination.

CONCLUSION

The OSCE appears to be a useful method for assessing final year osteopathy students undertaking various components of the typical osteopathic consultation. Having said this, it can still be argued that the validity of such an exam is reduced given that it does not faithfully replicate the osteopathic consultation in its entirety but breaks it into its separate components. This issue may be overcome when the results from the OSCE are combined with other forms of assessment (i.e. written assessments, portfolio) to develop a complete picture of the students fitness-to-practice as an osteopath. Modifications to the OSCE format have already been made for the examination to be conducted in 2012 and these changes will be the subject of further analysis.

REFERENCES

1. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990;**65**:910-7.
2. Rushforth HE. Objective structured clinical examination (OSCE): Review of literature and implications for nursing education. *Nurse Educ Today* 2007;**27**:481-90.
3. Schuwirth LWT, van der Vleuten CPM. The use of clinical simulations in assessment. *Med Educ* 2003;**37**:65-71.
4. Wass V, van der Vleuten CPM, Shatzer J, Jones R. Assessment of clinical competence. *Lancet* 2001;**357**:945-9.
5. McKnight J, Rideout E, Brown B, Cileska D, Patton D, Rankin J, et al. The objective structured clinical examination: an alternative approach to assessing student clinical performance. *J Nurs Educ* 1987;**26**:39-41.
6. Adome RO, Kitutu F. Creating an OSCE/OSPE in a resource-limited setting. *Med Educ* 2008;**42**:525-6.
7. Aeder L, Altshuler L, Kachur E, Barrett S, Hilfer A, Koepfer S, et al. The "culture OSCE": Introducing a formative assessment into a postgraduate program. *Educ Health* 2007;**20**:11p.
8. Adamo G. Simulated and standardized patients in OSCEs: Achievements and challenges 1992-2003. *Med Teach* 2003;**25**:262-70.
9. Schwartz RW, Witzke DB, Donnelly MB, Stratton T, Blue AV, Sloan DA. Assessing residents' clinical performance: Cumulative results of a four-year study with the Objective Structured Clinical Examination. *Surg* 1998;**124**:307-12.

10. Gleeson F. AMEE Medical Education Guide No. 9. Assessment of clinical competence using the Objective Structured Long Examination Recorder (OSLER). *Med Teach* 1997;**19**:7-14.
11. Norcini J. The death of the long case? *Br Med J* 2002;**324**:408.
12. Brenner AM. Uses and limitations of simulated patients in psychiatric education. *Acad Psych* 2009;**33**:112-9.
13. Bartfay WJ, Rombough R, Howse E, Leblanc R. The OSCE approach in nursing education... objective structured clinical examination. *Can Nurse* 2004;**100**:18-23.
14. Holmboe ES, Hawkins RE. Methods for evaluating the clinical competence of residents in internal medicine: A review. *Ann Intern Med* 1998;**129**:42-8.
15. Hassell AB. Assessment of specialist registrars in rheumatology: Experience of an objective structured clinical examination (OSCE). *Rheumatol* 2002;**41**:1323-8.
16. Olson LG. The ability of a long-case assessment in one discipline to predict students' performances on long-case assessments in other disciplines. *Acad Med* 1999;**74**:835-9.
17. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *Br Med J* 1975;**1**:447-51.
18. McCrorie P, Boursicot K. Variations in medical school graduating examinations in the United Kingdom: Are clinical competence standards comparable? *Med Teach* 2009;**31**:223-9.
19. Walsh M, Bailey PH, Koren I. Objective structured clinical evaluation of clinical competence: an integrative review. *J Adv Nurs* 2009;**65**:1584-95.
20. Sadler DR. Indeterminacy in the use of preset criteria for assessment and grading. *Assess Eval Higher Educ* 2009;**34**:159-79.
21. Hodges B, McIlroy JH. Analytic global OSCE ratings are sensitive to level of training. *Med Educ* 2003;**37**:1012-4.

22. Quirk M, Mazor K, Haley H-L, Wellman S, Keller D, Hatem D, et al. Reliability and validity of checklists and global ratings by standardized students, trained raters, and faculty raters in an objective structured teaching environment. *Teach Learn Med* 2005;**17**:202-9.
23. Al-Jarallah KF, Moussa MAA, Shehab D, Abdella N. Use of interaction cards to evaluate clinical performance. *Med Teach* 2005;**27**:369-74.
24. Brown G, Manogue M, Martin M. The validity and reliability of an OSCE in dentistry. *Eur J Dent Educ* 1999;**3**:117-25.
25. Khattab AD, Rawlings B. Assessing nurse practitioner students using a modified objective structured clinical examination (OSCE). *Nurse Educ Today* 2001;**21**:541-50.
26. Khattab AD, Rawlings B. Use of a modified OSCE to assess nurse practitioner students. *Br J Nurs* 2008;**17**:754-9.
27. Nayer M. An overview of the objective structured clinical examination. *Physiothera Canada* 1993;**45**:171-8.
28. Lindsey AA, Stritter F. Use of an OSCE in education for radiation therapists. *Radiol Technol* 1990;**62**:36-41.
29. Vaughan B, Sullivan V, Gosling C, McLaughlin P, Fryer G, Wolff M, et al. Methods of assessment used by osteopathic educational institutions. *Int J Osteopath Med* 2012;**15**:134-51.
30. London S. The assessment of clinical practice in osteopathic education: Is there a need to define a gold standard? *Int J Osteopath Med* 2008;**11**:132-6.
31. Abbey H. Assessing clinical competence in osteopathic education: Analysis of outcomes of different assessment strategies at the British School of Osteopathy. *Int J Osteopath Med* 2008;**11**:125-31.
32. Fletcher P. Clinical competence examination -- improvement of validity and reliability. *Int J Osteopath Med* 2008;**11**:137-41.

33. Baker HH, Cope MK, Adelman MD, Schuler S, Foster RW, Gimpel JR. Relationships between scores on the COMLEX-USA level 2-performance evaluation and selected [corrected] [published erratum appears in 2008 Apr;108(4):217]. *J Am Osteopath Assoc* 2006;**106**:290.
34. Cope MK, Baker HH, Foster RW, Boisvert CS. Relationships between clinical rotation subscores, COMLEX-USA examination results, and school-based performance measures. *J Am Osteopath Assoc* 2007;**107**:502-10.
35. Evans P, Goodson LB, Schoffman SI, Baker HH. Relations between academic performance by medical students and COMPLEX-USA Level 2: a multisite analysis. *J Am Osteopath Assoc* 2003;**103**:551-6.
36. Sevensma SC, Navarre G, Richards RK. COMLEX-USA and in-service examination scores: tools for evaluating medical knowledge among residents. *J Am Osteopath Assoc* 2008;**108**:713-6.
37. Fuller R, Pell G, Homer M, Roberts T. Comments on 'A systematic review of the reliability of objective structured clinical examination scores'. *Med Educ* 2012;**46**:337.
38. Pell G, Fuller R, Homer M, Roberts T. How to measure the quality of the OSCE: a review of metrics - AMEE guide no. 49. *Med Teach* 2010;**32**:802-11.

ACKNOWLEDGEMENTS

Chris Fallu for assistance with entering the data and for his assistance in ensuring the efficient conduct of the OSCE on the two days of the examination. Associate Professor Neil Diamond for assistance with the statistical analysis. Emily Jones (working with BV) developed the cases used for the OSCE and her time developing the cases and working on the administration of the examination is appreciated.

Student Name: _____ Examiner Name: _____

2011 Year 5 OSCE – Osteopathic Examination

Marking Criteria

1 = Below expected level of performance for Year 5 (unsatisfactory)

2 = Minimum/borderline level of performance for Year 5

3 = Expected level of performance for Year 5

4 = Above expected level of performance for Year 5

Identifies appropriate list of 3 most likely diagnoses.	4	3	2	1
Selects appropriate osteopathic examination for the patient's condition and the differential diagnoses, including passive, active and active resisted movements, taking into account patient factors such as pain levels, size, and available movement capacity.	4	3	2	1
Performs appropriate osteopathic examination competently.	4	3	2	1
Selects appropriate special/orthopaedic tests for patient's condition and the differential diagnoses.	4	3	2	1
Performs special/orthopaedic tests competently and interprets results accurately.	4	3	2	1
Explains the relevance of examinations to the patient and gains informed consent to perform them.	4	3	2	1
Ensures patient comfort throughout the examination.	4	3	2	1
Performs an examination that is in keeping with osteopathic principles.	4	3	2	1
Overall Performance (How did this student perform overall?)	S		U	

Figure 1. Marking sheet for the Osteopathic Examination station.

Station	Description
Case History	This station is designed to assess the student on their ability to interpret a case history, develop differential diagnoses, identify red and yellow flags as well as provide information to the examiners with elements of the patients' presentation that require further investigation or referral.
Clinical Diagnosis & Management	This station is designed to assess the student's ability to identify appropriate clinical examinations (neurological, cardiovascular, respiratory, abdomen) to assist the student with ruling in/out differentials and also identify when further referral is necessary.
Radiology	This station is designed to assess the student's ability to interpret a range of diagnostic imaging (plain film, CT, MRI) using a systematic process.
Pharmacology	This station is designed to assess the student's ability to identify medicines (both non-prescription and prescription) and demonstrate an understanding of the side effects, indications and contraindication of these medicines. In addition students are expected to discuss the impact of these medicines on the osteopathic management of the patient.
Patient Management	This station is designed to assess the student's ability to identify the issues that can impact on the overall management of the patient, particularly from a biopsychosocial viewpoint. Communication with the patient is also assessed.
Osteopathic Examination	This station is designed to assess the selection and performance of relevant osteopathic and orthopaedic examination procedures (in relation to the clinical history) in keeping with osteopathic principles.
HVLA	This station is designed to assess the student's consideration of safety related to the application of high velocity, low amplitude manipulation techniques. It also assesses the application of these techniques relevant to the clinical history and examination findings provided.
After-care	This station is designed to assess the ability of the student to provide focused, relevant, clinically indicated advice (lifestyle, nutrition, ergonomics) to the patient as part of the management plan. Advice provided should be based on the best available evidence.
Rehabilitation	This station is designed to assess the student's ability to incorporate appropriate and clinically relevant rehabilitation into the management plan for the patient. Advice provided should be based on the best available evidence.

	Panel 1		Panel 2		
	Average score	Number of fails	Average score	Number of fails	Number of fails (%)
Case History (20)	13.5 (3)	4	14 (2.5)	3	7 (14.5%)
Clinical Diagnosis & Management (24)	17 (4)	5	16 (3)	9	14 (30%)
Radiology (20)	14 (3)	2	14.5 (3)	3	5 (10.5%)
Pharmacology (16)	12 (2.5)	3	12 (2)	1	4 (8%)
Patient Management (24)	18 (3.5)	3	17.5 (3.5)	1	4 (8%)
Osteopathic Examination (32)	25 (3)	1	25 (3)	2	3 (6%)
HVLA (28)	20 (3)	2	20 (2.5)	3	5 (10.5%)
After-care (20)	15 (3)	5	14.5 (3.5)	0	5 (10.5%)
Rehabilitation (20)	14 (2.5)	0	15 (2)	3	3 (6%)

Table 2. Station name (total possible marks). Average score per station (standard deviation) and number of students who failed each station by panel and the per-station total.

	Examiner 1		Examiner 2	
	Pearson's <i>r</i>	R ²	Pearson's <i>r</i>	R ²
Case History	-0.55	0.30	-0.60	0.36
Clinical Diagnosis & Management	-0.55	0.30	-0.77	0.60
Radiology	-0.60	0.36	-0.58	0.33
Pharmacology	-0.64	0.40	-0.57	0.32
Patient Management	-0.44	0.19	-0.40	0.16
Osteopathic Examination	-0.63	0.40	-0.61	0.37
HVLA	-0.64	0.40	-0.60	0.36
After-care	-0.57	0.32	-0.59	0.35
Rehabilitation	-0.50	0.25	-0.46	0.21

Table 3. Pearson's *r* and R² coefficients for each station.

Table 4. Internal consistency measurements for each station. Value for each criteria represents the Cronbach's alpha if the criteria is deleted.

Station	Cronbach's alpha	Criteria 1	Criteria 2	Criteria 3	Criteria 4	Criteria 5	Criteria 6	Criteria 7	Criteria 8
Case history	0.81	0.76	0.80	0.76	0.74	0.81			
Clinical Diagnosis & Management	0.89	0.89	0.86	0.85	0.87	0.88	0.85		
Radiology	0.94	0.93	0.92	0.92	0.92	0.92			
Pharmacology	0.87	0.80	0.82	0.80	0.89*				
Patient Management	0.90	0.87	0.87	0.87	0.88	0.89	0.88		
Osteopathic Examination	0.82	0.82	0.78	0.79	0.80	0.81	0.81	0.81	0.82
HVLA	0.82	0.80	0.81	0.81	0.79	0.79	0.77	0.80	
After-care	0.91	0.88	0.89	0.88	0.90	0.88			
Rehabilitation	0.81	0.86*	0.73	0.74	0.73	0.78			

* *If item deleted* is greater than Cronbach's alpha.

Table 5. Variance components for each station.

Station	Student (s)	Case (c)	Examiner (e)	Student x Examiner	Case x Examiner
Case history	67%	27%		3%	3%
Clinical Diagnosis & Management	66%		17%	13%	4%
Radiology	61%	5%	10%	24%	
Pharmacology	63%		9%	20%	8%
Patient Management	70%		16%	3%	11%
Osteopathic Examination	86%			14%	
HVLA	50%			36%	14%
After-care	67%	2%		15%	16%
Rehabilitation	52%		5%	33%	10%

Note: Where there is no variance component listed, this indicates the facet is consistent across the student's total score for that station.

Table 6. Kappa scores for the global ratings each station.

Station	Kappa
Case History	0.80
Clinical Diagnosis & Management	0.90
Radiology	1.00
Pharmacology	0.64
Patient Management	1.00
Osteopathic Examination	1.00
HVLA	0.77
After-care	0.88
Rehabilitation	0.79

AUTHOR CONTRIBUTION STATEMENT

BV and PF developed the idea for the study and developed the method. BV and PF undertook the data analysis and developed the manuscript. All authors approved the final manuscript.

STATEMENT OF COMPETING INTERESTS

Brett Vaughan is an Editor of the Int J Osteopath Med but was not involved in review or editorial decisions regarding this manuscript.