Canadian Medical Education Journal
Revue canadienne de l'éducation médicale

What do we know about Objective Structured Clinical Examination in Sport and Exercise Medicine? A scoping review

Que savons-nous de l'examen clinique objectif structuré en médecine du sport et de l'exercice ? Une revue de l'étendue des connaissances

Reem El Sherif, Ian Shrier, Pierre-Paul Tellier and Charo Rodriguez

Volume 15, Number 3, 2024

URI: https://id.erudit.org/iderudit/1112775ar
DOI: https://doi.org/10.36834/cmej.77841

See table of contents

Publisher(s)

Canadian Medical Education Journal

ISSN

1923-1202 (digital)

Explore this journal

Cite this document

El Sherif, R., Shrier, I., Tellier, P.-P. & Rodriguez, C. (2024). What do we know about Objective Structured Clinical Examination in Sport and Exercise Medicine? A scoping review. *Canadian Medical Education Journal / Revue canadienne de l'éducation médicale*, *15*(3), 57–72. https://doi.org/10.36834/cmej.77841

Article abstract

Background and objectives: Despite the importance of the Objective Structured Clinical Examination (OSCE) in Sport and Exercise Medicine, the literature on the topic is fragmented and has been poorly developed. The goal of this review was to map current knowledge about how the OSCE is used in Sport and Exercise Medicine, and to identify knowledge gaps for future research.

Method: The authors conducted a scoping review. They searched PubMed and Scopus for articles using key terms related to 'OSCE' and 'sport medicine' with no limit on search start date and up to July 2022. Retrieved records were imported, abstracts were screened, and full-text articles were reviewed. A forward and backward citation tracking was conducted. Data was extracted and a qualitative meta-summary of the studies was conducted.

Results: A total of 469 records were screened, and 22 studies were included. The objectives of the studies included using OSCEs to assess knowledge/skills after a training program (n = 11), to assess an intervention (n = 8), and to assess and improve the OSCE itself (n = 3). Thirteen studies reported validity and/or reliability of the OSCE.

Conclusion: Despite the widespread use of OSCEs in the examination of Sport and Exercise Medicine trainees, only a handful of scholarly works have been published. More research is needed to support the use of OSCE in Sport and Exercise Medicine for its initial purpose. We highlight avenues for future research such as assessing the need for a deeper exploration of the relationship between candidate characteristics and OSCE scores.

Canadian Medical Education Journal

Reviews, Theoretical Papers, and Meta-Analyses

What do we know about Objective Structured Clinical Examination in Sport and Exercise Medicine? A scoping review

Que savons-nous de l'examen clinique objectif structuré en médecine du sport et de l'exercice ? Une revue de l'étendue des connaissances

Reem El Sherif,[1] Ian Shrier,[1,2] Pierre-Paul Tellier,[1] Charo Rodriguez[1]

[1]Department of Family Medicine, McGill University, Quebec, Canada; [2]Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, Montreal, Quebec, Canada

Correspondence to: Charo Rodriguez, MD, MSc, PhD, Professor of Family Medicine, McGill University, Director, McGill Family Medicine Education Research Group (FMER) ,5858 Côte-des-Neiges Road, 3rd floor, suite 300, room 328, Montreal, Quebec, Canada H3S 1Z1; phone: 1-514-399-9102; fax: 1-514-398-4202; email: charo.rodriguez@mcgill.ca

Published ahead of issue: Jun 24, 2024; published: Jul 12, 2024. CMEJ 2024, 15(3) Available at https://doi.org/10.36834/cmej.77841

Abstract

Background and objectives: Despite the importance of the Objective Structured Clinical Examination (OSCE) in Sport and Exercise Medicine, the literature on the topic is fragmented and has been poorly developed. The goal of this review was to map current knowledge about how the OSCE is used in Sport and Exercise Medicine, and to identify knowledge gaps for future research.

Method: The authors conducted a scoping review. They searched PubMed and Scopus for articles using key terms related to 'OSCE' and 'sport medicine' with no limit on search start date and up to July 2022. Retrieved records were imported, abstracts were screened, and full-text articles were reviewed. A forward and backward citation tracking was conducted. Data was extracted and a qualitative meta-summary of the studies was conducted.

Results: A total of 469 records were screened, and 22 studies were included. The objectives of the studies included using OSCEs to assess knowledge/skills after a training program (*n* = 11), to assess an intervention (*n* = 8), and to assess and improve the OSCE itself (*n* = 3). Thirteen studies reported validity and/or reliability of the OSCE.

Conclusion: Despite the widespread use of OSCEs in the examination of Sport and Exercise Medicine trainees, only a handful of scholarly works have been published. More research is needed to support the use of OSCE in Sport and Exercise Medicine for its initial purpose. We highlight avenues for future research such as assessing the need for a deeper exploration of the relationship between candidate characteristics and OSCE scores.

Résumé

Contexte et objectifs : Malgré l'importance de l'examen clinique objectif structuré (ECOS) en médecine du sport et de l'exercice, la littérature sur le sujet est fragmentée et peu développée. L'objectif de cette étude était de cartographier les connaissances actuelles sur l'utilisation de l'ECOS en médecine du sport et de l'exercice, et d'identifier les lacunes en matière de connaissances en vue de recherches futures.

Méthode : Les auteurs ont procédé à un examen approfondi. Ils ont recherché dans PubMed et Scopus des articles utilisant des termes clés liés à "OSCE" et "médecine du sport" sans limite de date de début de recherche et jusqu'en juillet 2022. Les enregistrements trouvés ont été importés, les résumés ont été examinés et les articles en texte intégral ont été examinés. Un suivi des citations en avant et en arrière a été effectué. Les données ont été extraites et un méta-résumé qualitatif des études a été réalisé.

Résultats : Au total, 469 dossiers ont été examinés et 22 études ont été incluses. Les objectifs des études comprenaient l'utilisation des ECOS pour évaluer les connaissances/compétences après un programme de formation (n = 11), pour évaluer une intervention (n = 8), et pour évaluer et améliorer l'ECOS lui-même (n = 3). Treize études ont fait état de la validité et/ou de la fiabilité des ECOS.

Conclusion : Malgré l'utilisation répandue des ECOS dans l'examen des stagiaires en médecine du sport et de l'exercice, seuls quelques travaux scientifiques ont été publiés. Des recherches supplémentaires sont nécessaires pour soutenir l'utilisation de l'OSCE en médecine du sport et de l'exercice pour son objectif initial. Nous mettons en évidence des pistes de recherche futures telles que l'évaluation de la nécessité d'une exploration plus approfondie de la relation entre les caractéristiques des candidats et les résultats des ECOS.

Introduction

The Objective Structured Clinical Examination (OSCE) employ standardised patients (SP) and stations to assess competency of candidates in clinical components such as history-taking, physical examination, knowledge application, clinical reasoning and communication skills.[1] Each component is separated into tasks corresponding to an objective checklist of items, used by the examiner (and at times the SP) to assess the candidates' performance. OSCEs are used as summative assessments, as well as for certification and licensure.[2] They are also used as formative tools which allow immediate observer feedback for the learner.[3]

The OSCE has become a customary method of assessment for undergraduate and postgraduate learners in many health disciplines and has widespread use in many countries.[4-7] Empirical studies and literature reviews on the topic of OSCEs either describe the results of the OSCEs in a specific department or medical field or focus on the performance of the OSCE itself.[8] The latter is assessed by examining the validity, reliability, cost-effectiveness, acceptability to candidates and evaluators, and setting of pass-fail standards.[9] Scholarly work on OSCEs has been abundant, covering different medical disciplines such as Family Medicine, Obstetrics and Gynecology, Orthopedics, and Psychiatry.[1,5,10-12] In some specialties where OSCEs are used, however, the research is less abundant. One of these specialties is Sport and Exercise Medicine.

Sport and Exercise Medicine

Sport and Exercise Medicine (SEM) is broadly defined as a specialty that encompasses "the variety of healthcare disciplines focused on physical activity-related performance and injury."[13] Some of the important skills needed for a sport medicine specialist include performing musculoskeletal (MSK) examination, diagnostic imaging, and diagnostic or therapeutic procedural skills, such as splints and casts.[14,15] As of 2009, there were 21 countries recognizing SEM as its own specialty, and another 15 countries that have sport medicine as a sub-speciality.[16] In most of these countries, practicing sport physicians have completed a postgraduate training that usually expands from two to five years.[17] In other words, SEM is now practiced at the specialist level in a substantial number of countries internationally, but that each country undertaking SEM specialist training had essentially "re-invented the wheel."[18]

In Canada, SEM is not considered its own specialty, and physicians come from different specialties including family medicine, orthopedic surgery, rheumatology and internal medicine.[15] After completing training in their specialty, some physicians complete additional specialized SEM training during a one-year fellowship organized by a Canadian University Faculty of Medicine program.[19]

OSCE of Sport and Exercise Medicine

Since 1989, the Credentials committee of the Canadian Academy of Sport and Exercise Medicine (CASEM) has annually administered an OSCE exam of SEM. The development, format, administration and scoring of this OSCE has been described elsewhere, and it has shown high reliability and validity.[20,21] The examination is currently open to any medical doctor of any specialty who has a minimum of two years of independent medical practice and an active license, and to any Canadian family medicine resident who is completing at least a one-year Sport Medicine fellowship recognized by a University Faculty of Medicine Program. The examination is now also used as an evaluation tool by Fellowship Directors for the sport medicine Certificate of Added Competence (CAC) created by the College of Family Physicians of Canada (CCFP(SEM)). [19] Additionally, as of July 2020, it is one of the requirements of the Standards of Accreditation for Areas of Focused Competence Programs in SEM by the Royal College of Physicians and Surgeons of Canada.[15]

Despite the importance of the OSCE in SEM and its usage in high-stakes licensing exams, the literature about the use of OSCEs in SEM is fragmented and has been poorly developed. This presents a barrier for researchers in medical education who wish to explore avenues for research into this topic, including us. Therefore, our review was guided by two objectives: (1) to map current literature about how and when OSCEs are used to assess medical trainees' (including students, residents, and licensed physicians) competency in SEM, and (2) to identify knowledge gaps for future research.

Methods

Research design and review question

We conducted a scoping review of the literature on OSCEs in SEM using the framework detailed by Arksey and O'Malley.[22] Unlike systematic reviews, scoping reviews address a broad research question and do not require an exhaustive search of the literature, two independent reviewers, or critical appraisal of included studies.[22-25] We followed the Preferred Reporting Items for Systematic

reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist.[26] The review process was guided by the following overarching question: What do we currently know about the use of OSCEs to assess medical trainees' (including students, residents, and licensed physicians) competency in SEM? Due to the nature of the review (not systematic) it was not registered in PROSPERO.

Identifying relevant studies: search strategy and information sources

A search strategy was compiled with the help of a health librarian. Key terms related to 'OSCE' and 'sports medicine' were searched in PubMed (including Medline) and Scopus. The full search strategy is presented in the Appendix. There was no limit on the language of publication, type of record or date of publication. Deduplicated records were exported into an online software for conducting reviews (Distiller SR) where the stages of selection and data extraction occurred.

After the selection stage, additional potentially relevant records were retrieved by forward and backward citation tracking in Scopus. The initial search was conducted from inception in July and August 2018, and an updated search was conducted in July 2022. The search criteria were not modified in the updated search.

Study selection and data synthesis

The retrieved abstracts and subsequent full-text articles were reviewed by a single author following a pilot testing of the criteria with a second author. The following inclusion/eligibility criteria were used: (1) the article is an empirical study or review; (2) the article is on OSCEs or similar assessments using stations with case-based scenarios;[12] and (3) the article is about sport/exercise/MSK medicine or the assessment of skills that are relevant to training in sport medicine. We excluded articles describing OSCEs used for assessment in general medical training or in other specialties.

Charting and synthesizing the data

The lead author imported the included studies into NVivo qualitative software and, in an iterative manner with the co-authors, synthesis was conducted using a qualitative meta-summary which describes the process of "grouping, abstraction, and formatting of findings, and the calculation of frequency."[27] The following data was coded in NVivo: country of study, study design, study objective, study participants, description of the SEM/MSK medicine training, OSCE description, study findings, and study limitations. We specifically focus on two types of evaluations of OSCE performance: reliability and validity. Reliability measurements include internal consistency, inter-rater, intra-rater and overall reliability. Validity measurements include construct validity, content validity, concurrent validity, and face validity. Although a more contemporary view of validity places previous distinctions of validity within construct validity, this review describes the distinct forms of validity as described by the authors of the included studies.[28] When validity measures were not directly stated by the authors, we inferred them from the description of the OSCE, e.g., using expert consensus to develop the rater checklist would account for content validity.

Results

We first screened 469 unique abstracts, and then reviewed 36 full-text articles, to end up with the selection of 23 original empirical studies that met the eligibility criteria. An additional study was excluded after discussion with the authors as it did not include medical trainees. Full details are in the PRISMA flow diagram in Figure 1.

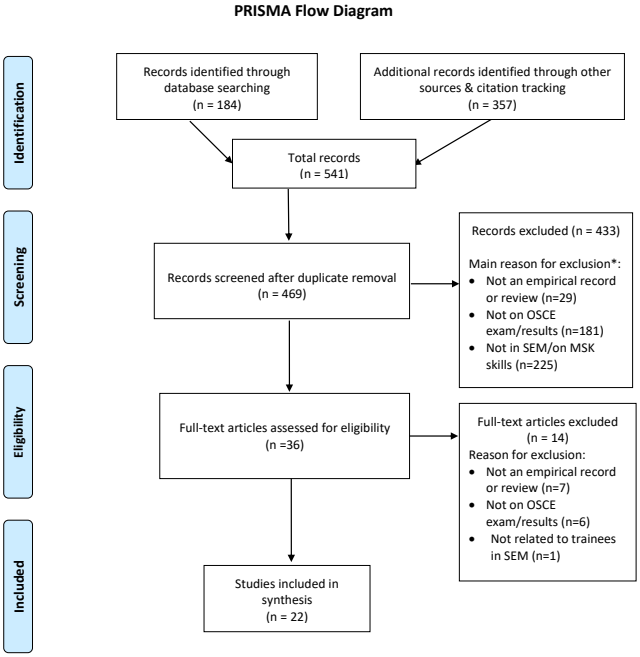

Figure 1. PRISMA flow diagram
**There may be more than one reason for exclusion.*

Of the 22 included studies, almost two thirds were descriptive, 77% were from North America, and more than half focused on residents and fellows. OSCEs on more general MSK examination were described in 61% of the studies and the number of OSCE stations study ranged from a single station to 20 stations. All the studies describe the OSCE itself, but the majority of studies (55%) did not discuss standard setting (methods of assigning pass/fail

scores), as pass scores were not relevant to their study purpose. More details are in Table 1 and full details on the included studies can be found in Appendix A.

Table 1. Characteristics of included studies

		N of articles (%)
Research design	single-group descriptive studies	14 (63.6)
	randomized controlled trials	5 (22.7)
	case-control studies	3 (13.6)
Country	Canada	9 (40.9)
	USA	9 (40.9)
	United Kingdom	1 (4.5)
	Australia	1 (4.5)
	China	1 (4.5)
	Iran	1 (4.5)
Participants	residents and fellows	13 (59)
	medical students	7 (31.8)
	practicing physicians	1 (4.5)
	department faculty	1 (4.5)
OSCE topic	General musculoskeletal (MSK) examination	13 (59)
	SEM	6 (27.2)
	Specific MSK skills	3 (13.6)
OSCE stations	>3 stations	10 (45.4)
	2-3 stations	6 (27.2)
	1 station	5 (22.7)
	Not mentioned	1 (4.5)
Standard setting	Not mentioned	12 (54.5)
	Described in detail	10 (45.4)
Study setting	Orthopedics or SEM department	9 (40.9)
	Medical school	6 (27.2)
	Physical Medicine and Rehabilitation	2 (9.1)
	Veterans Affairs medical centers	1 (4.5)
	Other departments	4 (18.2)
Objective	Assessment of participant knowledge and skills during or after a training program	11 (50)
	Assessment of an educational intervention	8 (36.3)
	Assessment and improvement of an OSCE tool	3 (13.6)

Studies that use an OSCE during or at the end of training to assess trainees' competency

Four studies were conducted by the same first author in a Canadian Orthopedics department and used an OSCE to assess residents or fellows after completing a rotation or fellowship year in SEM. In the first study, the authors developed and assessed the performance of a SEM OSCE, and reported an overall Cronbach alpha of 0.91 and the ability of the OSCE to discriminate among different years of training.[29] This OSCE was deemed of high reliability and validity and, because the authors had access to the data, it was used in the next three studies. One study reported that the OSCE scores of senior residents were consistently higher than junior residents: further confirming the construct validity of the OSCE in this setting.[30] In the other two studies, the authors used the OSCE to assess clinical skills and other tools to assess technical skills, reporting moderate concurrent validity based on the correlation between both sets of scores.[31,32]

In most of the remaining studies, the authors report that OSCE scores demonstrated that candidates' skills improved following their training.[33-36] However, one study used an OSCE to assess Orthopedics residents' skills during various stages of their training and reported a relatively low average score (66%).[37] Another study involved second-year medical students and described the design, development, and use of four checklists for evaluating MSK examination. The authors report high reliability of the checklists, yet the students' scores ranged from 22% to 50%.[38] A recent study compared the performance of nine orthopedics residents in competency-based curriculum exams that included OSCEs, and the multiple choice, computer-based Orthopaedic In-Training Examination (OITE) over two years.[39] The authors reported a significant improvement in mean OITE scores between the two years, but no significant correlation between OITE improvement and OSCE scores.

Table 2. Reporting OSCE validity and reliability

Study	OSCE validity	OSCE reliability
Battistone 2016[33]- 1wk	Previously validated in other studies but did not provide further details	Previously validated in other studies: Interrater reliability was 87% (k ¼ 0.61) and 97% (k ¼ 0.88) for the knee and shoulder, respectively.
Battistone et al 2016[34] - 3day	Previously validated in other studies but did not provide further details	Previously validated in other studies interrater reliability was 87% (k ¼ 0.61) and 97% (k ¼ 0.88) for the knee and shoulder, respectively.
Beran 2012[37]	Not reported	Interrater reliability: Averaging the scores from the three raters improved the reliability to almost perfect for all intraclass correlation coefficient values except that for the trauma PE, which improved from fair to moderate.
Chen 2013[48]	Construct validity: high Cronbach's α for the23-item scale regarding global assessment in all 3 tests (0.932 for the pretest, 0.926 for the post test, and 0.892 for the long-term test). Expert validity: p-value was 0.626 (independent t-test, n.s).	Kendall's W showed only moderate inter-rater reliability. KR-20 was 0.96 for the pretest, 0.968 for the post test, and 0.892 for the long-term test, indicating high internal consistency.
Dwyer 2013[29]	OSCE questions and binary checklists underwent formal content review at a series of focus groups of orthopaedic surgeons to establish content validity. Concurrent validity by comparing to OITE scores showed high correlation.	Internal consistency reliability: Cronbach α was 0.91 for the six stations. Overall reliability was also high
Dwyer, Wright, et al., 2016[49]	*similar to Dwyer 2013	*similar to Dwyer 2013
Dwyer 2015[30]	*similar to Dwyer 2013	*similar to Dwyer 2013
Dwyer 2020[32]	*similar to Dwyer 2013 Concurrent validity by comparing to scores in technical skills assessment tools showed moderate correlation.	Internal consistency reliability: Cronbach α for the entry OSCE was 0.88 and for the exit OSCE it was 0.81.
Dwyer, Slade et al., 2016[31]	Content validity: validated checklist and development of checklist by expert consensus. Construct validity: there was evidence of novice-expert differentiation. Concurrent validity: a good correlation between performance on the OSATS and the end of rotation OSCE.	Excellent internal consistency. The overall reliability of the OSATS (0.9) and the inter-rater reliability (0.9) were both high.
Harasym 1997[21]	The OSCE enjoys high face validity and this study provided evidence for construct validity	Highly reliable (0.83-0.95) in other studies.
Irwin 2018[36]	Not reported	Not reported
Jain 1997[35]	Not reported	Not reported
McGaghie 1994[38]	Not reported	Interrater reliability was assessed for the data from each checklist for each year using weighted Kappa. The data show that, with one exception interrater reliability was high. The internal consistency reliability of the checklist data was estimated using Cronbach's α. The data show uniformly high alpha coefficients ranging from 0.72 to 0.92.
Mehrpour 2013[44]	Not reported	Internal consistency reliability of the 13-item OSCE overall performance scale was acceptable and yielded a Cronbach's α of 0.91.
Okoro 2021[39]	Not reported	Not reported
Oswald 2011[40]	Not reported	Internal consistency reliability for the five station OSCE was α of 0.67
Raj 2006[41]	Previously validated in other studies but did not provide further details	Not reported
Siddharthan 2017[45]	Not reported	Not reported
Smith 2002[46]	Not reported	Not reported
Smith 2005[43]	Not reported	Not reported (To ensure reliability, the pre- and post-tests were identical, and one of two faculty members was present at every OSCE station and completed the checklist during the assessment period.
Smith 2000[42]	Not reported	Not reported
Yu 2020[47]	Not reported	Not reported

Studies that use an OSCE to assess the effectiveness of an educational intervention

Three studies compared the OSCE scores of candidates taught by patients trained in MSK examination skills (patient educators) with candidates taught by physicians. All three studies reported no significant difference in OSCE scores between the two groups.[40-42] Five remaining studies explored additional MSK training in the form of lectures, a workshop, a supplemental instructional video, spaced education or structured clinical instruction modules. All studies reported that candidates exposed to training scored higher or had similar OSCE scores compared to the control group.[43-47]

OSCE to assess and improve an OSCE tool

Three studies focused on assessing and improving the OSCE itself. The oldest study examined the relationship of the CASEM-OSCE scores and expertise in the field to assess the exam's construct validity.[21] The study reported a significant positive relationship between OSCE scores and expertise level, providing evidence of the exam's construct validity.

One study focused on the OSCE examiners rather than the candidates. It explored the use of a videotaped OSCE in a fracture scenario to assess internal consistency and inter-rater reliability of raters from different subspecialties and with varying levels of seniority.[48] The raters used validated checklists to assess the residents' performance at three time points. Calculations revealed high construct validity, moderate inter-rater reliability, and high internal consistency. These findings suggest that the rater does not necessarily have to be a specific person and this finding make an OSCE more accessible and practical to run. These results also demonstrate that using OSCE recordings was an acceptable method for future evaluator training and for credentialing purposes.

The third study compared three different methods for standard setting and reported that the Modified Angoff method was an acceptable method of setting different pass marks for senior and junior residents.[49] This finding enables both senior and junior residents to sit the same OSCE, thus reducing the burden of developing and conducting OSCEs. The authors also reported that the pass marks were not significantly changed by the subspecialty training of the OSCE judges.

Reporting OSCE validity and reliability

Validity of the OSCE was reported in nine studies and reliability of the OSCE was reported in 13 studies (Table 2). Four studies by the same author used the same OSCE questions and checklist with established content and concurrent validity, as well as reliability.[29,30,32,49] Concurrent validity was assessed by comparing the study OSCE to OITE scores. The fifth study by the same author reports content, construct and concurrent validity of the OSCE method they developed.[31] Three studies used checklists previously validated in other studies but did not provide further details.[33,34,41] Ten studies reported high internal consistency reliability using Cronbach's alpha with values ranging from 0.67 to 0.91. Only seven studies reported interrater reliability. One study did not assess reliability but reported measures to ensure reliability including having one of two faculty members present at every OSCE station.[43] One study did not describe the OSCEs used in any detail.[39]

Discussion

This scoping review first unveils the low prevalence of published works in which OSCEs were used for assessing the competency of SEM trainees. In fact, only six out of the 22 studies included in this review were developed in a SEM setting, and five of those were conducted by the same author in the same setting.[29-32,49] It is implausible that fewer OSCEs are being conducted in Sport Medicine over time; the OSCE has been conducted by CASEM since 1989 and has been used by different medical schools across the UK and Australia since at least 2000.[50-52] The American Medical Society for Sports Medicine has proposed using the OSCE as one of the methods to directly observe and document fellows' skills in Sports Medicine fellowship programs.[53] It is, thus, more likely that few medical education scholars are doing research on this topic or that many assessments of training programmes are not yet published. This may be an area for future investigation.

This scoping review also unveils that OSCEs were generally used for two different objectives in Sports Medicine and MSK education programs: (1) to assess trainees' competency in this medical discipline following a training program, and (2) to assess the effectiveness of MSK or Sports Medicine education interventions. A large number of studies in our review reported measures of validity and reliability, which are two of the important measures when OSCEs are used for summative or certification purposes.

On the other hand, the use of the OSCE is not exempt of criticism. A number of scholars claim that dichotomous checklists actually reduce the ability to meaningfully evaluate clinical competence,[54] and highlight the weaknesses of standardizing the assessment of health

professional trainees.[55] However, only one study in our review explored this controversial view by comparing the OSCE to another examination that allowed for more rater flexibility. [56] The five studies conducted by Dwyer et al. assessed concurrent validity by comparing scores on OSCEs to scores of other evaluation tools such as the OITE and exams assessing technical skills. Two other studies aimed to assess and improve the OSCE tool itself by comparing two different methods of conducting it. Another study, focusing on the CASEM-OSCE, assessed construct validity by comparing OSCE scores to candidate's expertise in SEM and reported a significant, positive relationship between both.[21]

Our second objective was to identify knowledge gaps and limitations in the research on Sport Medicine OSCEs to suggest future research avenues. In this regard, our review uncovered the need to improve research methods to adequately evaluate the use of OSCEs in SEM. The limitation that was most commonly reported by the authors of the included studies was the relatively small sample size. Only one study in our review included the results of the SEM OSCEs longitudinally to assess the OSCE's validity,[21] while two studies compared results from two years.[39,46] In future research, such data would allow researchers to conduct an historical longitudinal study to assess re-failure rates, the relationship between specific candidate characteristics and OSCE scores, how individual examiners may change over time, and so on. Moreover, studies conducted on this topic should more thoroughly involve appropriate measurements of assessment quality such as utility, acceptability, and validity.

This review explored OSCEs in SEM.. One systematic review in Family Medicine focused on documenting the quality of SEM OSCEs in terms of psychometrics and standard setting, and reported that few articles reported the psychometric methods and results.[8] In fact, another review identified the need for standardized reporting within OSCE studies in medical education and provided a checklist to guide future researchers.[57] Other reviews focused on specific aspects of the OSCE such as communication skills,[58,59] the role of peer-assessment[60] and the association between test anxiety and OSCE performance.[61] It is unclear how findings about OSCEs from other specialties and various aspects of performance are relevant to SEM OSCEs.

Limitations

Our search strategy focused on terms related to OSCEs and may have missed other examinations that use case-based scenarios and checklists. We tried to overcome this by using the MeSH term "educational measurement/methods" to retrieve some of these other studies. We were unable to perform any quantitative analysis on the results of the included studies due to the diversity of study designs and objectives. Finally, while we selected studies were the OSCEs described were relevant to SEM, the participants of the included studies were largely not SEM physicians. Future work should explore how the SEM competencies diverge and converge from other specialties, such as between Orthopedic Surgery and Family Medicine.

Conclusion

OSCEs are generally used for two purposes in this field: to assess trainee competence following a training program, and to assess the effectiveness of educational interventions. Our review has revealed that, despite the widespread use of OSCEs in the examination of SEM trainees, little is known about this use because only a handful of scholarly works has been published. There remains a strong need to improve research methods to adequately evaluate the use of OSCEs in SEM, specifically by increasing the sample size of studies and ensuring studies that are done in this field involve appropriate measurements of assessment.

Conflicts of Interest: None

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Authorship: All the authors were responsible for developing the research design and search strategy. RES performed the literature search, selection, data extraction and qualitative meta-summary while completing her PhD program. All authors contributed to the iterative drafting and refinement of the manuscript and approved the final version of the manuscript for submission.

Acknowledgement: We would like to acknowledge Genevieve Gore, health librarian Schulich Library of Physical Sciences, Life Sciences, and Engineering at McGill University, for her help in guiding the search strategy.

Edited by: Marco Zaccagnini (senior section editor); Marcel D'Eon (editor-in-chief)

References

1. Kreptul D, Thomas R. Family medicine resident OSCEs: a systematic review. *Education for primary care: an official publication of the Association of Course Organisers, National Association of GP Tutors, World Organisation of Family Doctors*. Nov 2016;27(6):471-477. https://doi.org/10.1080/14739879.2016.1205835
2. Lafave M, Mohtadi N, Chan D. Musculoskeletal injury evaluation standards for different disciplines. *Athletic Therapy Today*. 2012;17(2):21-24. https://doi.org/10.1123/ijatt.17.2.21
3. Martinez J, Harris C, Jalali C, Tung J, Meyer R. Using peer-assisted learning to teach and evaluate residents' musculoskeletal skills. *Med Educ Online*. 2015;20(1)27255. https://doi.org/10.3402/meo.v20.27255
4. Harden RM. Revisiting 'Assessment of clinical competence using an objective structured clinical examination (OSCE)' *Med Educ*. 2016;50(4):376-379. https://doi.org/10.1111/medu.12801
5. Hodges B, Hollenberg E, McNaughton N, Hanson M, Regehr G. The psychiatry OSCE: a 20-year retrospective. *Acad Psychiatry*. 2014;38(1):26-34. https://doi.org/10.1007/s40596-013-0012-8
6. Zayyan M. Objective structured clinical examination: the assessment of choice. *Oman Med J*. 2011;26(4):219. https://doi.org/10.5001/omj.2011.55
7. Sloan DA, Donnelly MB, Schwartz RW, Strodel WE. The objective structured clinical examination: the new gold standard for evaluating postgraduate clinical performance. *Annals Surg*. 1995;222(6):735-742. https://doi.org/10.1097/00000658-199512000-00007
8. Kreptul D, Thomas R. Family medicine resident OSCEs: a systematic review. *Educ Prim Care*. 2016;27(6):471-477. https://doi.org/10.1080/14739879.2016.1205835
9. Hodges B, Hanson M, McNaughton N, Regehr G. Creating, monitoring, and improving a psychiatry OSCE. *Acad Psychiatry*. 2002;26(3):134-161. https://doi.org/10.1176/appi.ap.26.3.134
10. Casey PM, Goepfert AR, Espey EL, et al. To the point: reviews in medical education-the Objective Structured Clinical Examination. *Amer J Obstetric Gynecol*. 2009;200(1):25-34. https://doi.org/10.1016/j.ajog.2008.09.878
11. McNaughton N, Ravitz P, Wadell A, Hodges BD. Psychiatric education and simulation: A review of the literature. *Can J Psychiatry*. 2008;53(2):85-93. https://doi.org/10.1177/070674370805300203
12. Dwyer T, Glover Takahashi S, Kennedy Hynes M, et al. How to assess communication, professionalism, collaboration and the other intrinsic CanMEDS roles in orthopedic residents: use of an objective structured clinical examination (OSCE). *Can J Surg*. 2014;57(4):230-6. https://doi.org/10.1503/cjs.018813
13. Safai P. A Critical Analysis of the Development of Sport Medicine in Canada, 1955–80. *International Review for the Sociology of Sport*. 2007;42(3):321-341. https://doi.org/10.1177/1012690207088115
14. Wesner M. What is a sport and exercise medicine physician? *Can Fam Phys*. 2017.
15. The Royal College of Physicians and Surgeons of Canada. *Competency training requirements for the area of focused competence in sport and exercise medicine.* http://www.royalcollege.ca/rcsite/documents/ibd/sport-exercise-medicine-ctr-e [Accessed Jun 14, 2020].
16. Pigozzi F. Specialisation in sports medicine: the state of the sport medicine specialty training core curriculum in the European Union. *Brit Assoc Sport Excercise Med*. 2009. https://doi.org/10.1136/bjsm.2008.055350
17. Orchard JJ, Orchard JW, Driscoll TR. Comparison of sports medicine, public health and exercise promotion between bidding countries for the FIFA World Cup in 2018. *Brit J Sports Med*. 2010;44(9):631-636. https://doi.org/10.1136/bjsm.2010.073551
18. Humphries D, Jaques R, Dijkstra HP. A Delphi developed syllabus for the medical specialty of sport and exercise medicine. *Brit J Sports Med*. 2018;52(8):490-492. https://doi.org/10.1136/bjsports-2017-098477
19. CASEM/ACMSE. *Sport and Exercise Medicine Diploma Exam.,* https://casem-acmse.org/education/diploma-in-sem/ [Accessed Jun 12, 2020].
20. Mohtadi NGH, Harasym PH, Pipe AL, Strother RT, Mah AF. Using an objective structured clinical examination to evaluate competency in sport medicine. *Clin J Sport Med*. 1995;5(2):82-85. https://doi.org/10.1097/00042752-199504000-00002
21. Harasym P, Mohtadi N, Henningsmoen H. Construct validity of “high stakes” OSCE scores. *Adv Med Educ*. Springer; 1997:443-445. https://doi.org/10.1007/978-94-011-4886-3_134
22. Arksey H, O'malley L. Scoping studies: towards a methodological framework. *Intern J Soc Res Meth*. 2005;8(1):19-32. https://doi.org/10.1080/1364557032000119616
23. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci*. 2010;5(1):69. https://doi.org/10.1186/1748-5908-5-69
24. Morris M, Boruff JT, Gore GC. Scoping reviews: establishing the role of the librarian. *JMLA*. 2016;104(4):346. https://doi.org/10.3163/1536-5050.104.4.020
25. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. *JBI Evidence Synthesis*. 2020;18(10):2119-2126. https://doi.org/10.11124/jbies-20-00167
26. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals Intern Med*. 2018;169(7):467-473. https://doi.org/10.7326/M18-0850
27. Sandelowski M, Barroso J, Voils CI. Using qualitative metasummary to synthesize qualitative and quantitative descriptive findings. *Res Nurs health*. 2007;30(1):99-111. https://doi.org/10.1002/nur.20176
28. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Amer J Med*. 2006;119(2):166.e7-166.e16. https://doi.org/10.1016/j.amjmed.2005.10.036
29. Dwyer T, Theodoropoulos J, Herold J, et al. Assessing competence of orthopaedic residents: the reliability and validity of an objective structured clinical examination after a sports medicine rotation. *J Bone Joint Surg - Series A*. 2013;95(22):e1771-e1779. https://doi.org/10.2106/JBJS.M.00148
30. Dwyer T, Wright S, Kulasegaram K, et al. Competency-based medical education can both junior residents and senior

residents achieve: competence after a sports medicine training module? *J Bone Joint Surg - Amer Vol*. 2015;97(23):1985-1991. https://doi.org/10.2106/JBJS.O.00252
31. Dwyer T, Slade Shantz J, Kulasegaram KM, et al. Use of an objective structured assessment of technical skill after a sports medicine rotation. *J Arthroscopy*. 2016;32(12):2572-2581.e3. https://doi.org/10.1016/j.arthro.2016.05.037
32. Dwyer T, Chahal J, Murnaghan L, et al. Development of a certification examination for orthopedic sports medicine fellows. *Can J Surg.* 2020;63(2):E110. https://doi.org/10.1503/cjs.015418
33. Battistone MJ, Barker AM, Grotzke MP, et al. Effectiveness of an interprofessional and multidisciplinary musculoskeletal training program. *J Grad Med Ed*. 2016;8(3):398-404. https://doi.org/10.4300/JGME-D-15-00391.1
34. Battistone MJ, Barker AM, Grotzke MP, Beck JP, Lawrence P, Cannon GW. "Mini-Residency" in musculoskeletal care: a national continuing professional development program for primary care providers. *J Gen Intern Med*. 2016;31(11):1301-1307. https://doi.org/10.1007/s11606-016-3773-4
35. Jain SS, Nadler S, Eyles M, Kirshblum S, Delisa JA, Smith A. Development of an objective structured clinical examination (OSCE) for physical medicine and rehabilitation residents. *Amer J Phys Med Rehab* 1997;76(2):102-106. https://doi.org/10.1097/00002060-199703000-00003
36. Irwin R, Smith J, Issenberg SB. Long-term retention of musculoskeletal ultrasound training during residency. *Amer J Phys Med Rehab*. 2018;97(7):523-530. https://doi.org/10.1097/PHM.0000000000000924
37. Beran MC, Awan H, Rowley D, Samora J, Griesser MJ, Bishop JY. Assessment of musculoskeletal physical examination skills and attitudes of orthopaedic residents. *J Bone Joint Surg - Series A*. 2012;94(6):e36(1). https://doi.org/10.2106/JBJS.K.00518
38. McGaghie WC, Renner BR, Kowlowitz V, et al. Development and evaluation of musculoskeletal performance measures for an objective structured clinical examination. *Teach Learn Med*. 1994;6(1):59-63. https://doi.org/10.1080/10401339409539645
39. Okoro T, Mironova P, Sathiyamoorthy T, et al. Correlation of residents' performance in competency-based exams and orthopaedic in-training examinations (OITE). 2021; https://doi.org/10.21203/rs.3.rs-426917/v1
40. Oswald A, Bell M, Wiseman J, Snell L. The impact of trained patient educators on musculoskeletal clinical skills attainment in pre-clerkship medical students. *BMC Med Educ*. 2011;11(1)65. https://doi.org/10.1186/1472-6920-11-65
41. Raj N, Badcock L, Brown G, Deighton C, O'Reilly S. Undergraduate musculoskeletal examination teaching by trained patient educators - a comparison with doctor-led teaching. *Rheumatol*. 2006;45(11):1404-1408. https://doi.org/10.1093/rheumatology/kel126
42. Smith MD, Henry-Edwards S, Shanahan EM, Ahern MJ. Evaluation of patient partners in the teaching of the musculoskeletal examination. *J Rheumatol.* 2000;27(6):1533-1537.
43. Smith C, Newman B, Davis R, Yang J, Ramanan R. A comprehensive new curriculum to teach and assess resident knowledge and diagnostic evaluation of musculoskeletal complaints. *Med Teach*. 2005;27(6):553-558. https://doi.org/10.1080/01421590500156152
44. Mehrpour SR, Aghamirsalim M, Motamedi SMK, Ardeshir Larijani F, Sorbi R. A supplemental video teaching tool enhances splinting skills basic research. *Clinic Orth Related Res*. 2013;471(2):649-654. https://doi.org/10.1007/s11999-012-2638-3
45. Siddharthan T, Soares S, Wang HH, Holt SR. Objective structured clinical examination-based teaching of the musculoskeletal examination. *Southern Med J*. 2017;110(12):761-764. https://doi.org/10.14423/SMJ.0000000000000739
46. Smith M, Walker J, Schultz D, et al. Teaching clinical skills in musculoskeletal medicine: the use of structured clinical instruction modules. *J Rheumatol*. Apr 2002;29(4):813-7.
47. Yu JC, Guo Q, Hodgson CS. Deconstructing the joint examination: a novel approach to teaching introductory musculoskeletal physical examination skills for medical students. *MedEdPORTAL*. 2020;16:10945-10945. https://doi.org/10.15766/mep_2374-8265.10945
48. Chen ACY, Lee MS, Chen WJ, Lee ST. Assessment in orthopedic training - an analysis of rating consistency by using an objective structured examination video. *J Surg Educ*. 2013;70(2):189-192. https://doi.org/10.1016/j.jsurg.2012.11.002
49. Dwyer T, Wright S, Kulasegaram KM, et al. How to set the bar in competency-based medical education: standard setting after an Objective Structured Clinical Examination (OSCE). *BMC Med Educ*. 2016;16(1)506. https://doi.org/10.1186/s12909-015-0506-z
50. Macleod DA. Intercollegiate board for sport and exercise medicine. *Brit J Sports Med*. 2000;34(4):235-235. https://doi.org/10.1136/bjsm.34.4.235
51. West LR, Griffin S. Sport and exercise medicine in the UK: what juniors should know to get ahead. 2017. *Brit J Sports Med.* https://doi.org/10.1136/bjsports-2016-096631.
52. Fricker P. Sports medicine education in Australia. *Brit J Sports Med*. 2000;34(4):240-241. https://doi.org/10.1136/bjsm.34.4.240
53. Asif IM, Stovak M, Ray T, Weiss-Kelly A. Primary care sports medicine fellowship: AMSSM proposed standards of excellence. *Clin J Sport Med*. 2017;27(3):231-244. https://doi.org/10.1097/JSM.0000000000000428
54. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. *Acad Med*. 1999;74(10):1129-1134. https://doi.org/10.1097/00001888-199910000-00017
55. Reid H, Gormley GJ, Dornan T, Johnston JL. Harnessing insights from an activity system–OSCEs past and present expanding future assessments. *Med Teach*. 2020:1-6. https://doi.org/10.1080/0142159X.2020.1795100
56. Lafave MR, Katz L. Validity and reliability of the standardized orthopedic assessment tool (SOAT): a variation of the traditional objective structured clinical examination. *J Athlet Train*. 2014;49(3):373-380. https://doi.org/10.4085/1062-6050-49.1.12
57. Patricio M, Juliao M, Fareleira F, Young M, Norman G, Vaz Carneiro A. A comprehensive checklist for reporting the use of

OSCEs. *Med Teach*. 2009;31(2):112-24. https://doi.org/10.1080/01421590802578277

58. Setyonugroho W, Kennedy KM, Kropmans TJB. Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: a systematic review. *J Pat Educ Counsel*. 2015. https://doi.org/10.1016/j.pec.2015.06.004
59. Comert M, Zill JM, Christalle E, Dirmaier J, Harter M, Scholl I. Assessing communication skills of medical students in objective structured clinical examinations (osce)--a systematic review of rating scales. *PloS one*. 2016;11(3):e0152717. https://doi.org/10.1371/journal.pone.0152717
60. Khan R, Payne MWC, Chahine S. Peer assessment in the objective structured clinical examination: a scoping review. *Med Teach*. 2017;39(7):745-756. https://doi.org/10.1080/0142159X.2017.1309375
61. Martin RD, Naziruddin Z. Systematic review of student anxiety and performance during objective structured clinical examinations. *Curr Pharm Teach Learn*. 2020;12(12):1491-1497. https://doi.org/10.1016/j.cptl.2020.07.007

Appendix A. Full details on the included studies

Table 3. Search strategy and information sources

Database	Date of the search	Concepts	Search strategy
SCOPUS	July 17th, 2018 Update: July 31st, 2022	OSCE and sport/ musculoskeletal/ exercise medicine	Objective Structured Clinical Examination OR “OSCE” AND “sport medicine” OR “musculoskeletal medicine” OR “exercise medicine”
PubMed (includes Medline)	August 5th, 2018 Update: July 31st, 2022	OSCE and sport medicine/exercise medicine	Search ((osce[tw] OR objective structured clinical exam*[tw] OR "educational measurement/methods"[mesh])) AND ("sport medicine"[mesh] OR sport medicine[tw] OR musculoskeletal medicine[tw] OR exercise medicine[tw])

Table 4. Characteristics of included studies

Study & country	Study design	Study objective	Participants	SEM/MSK training	OSCE topic & number of stations	Results/effect of intervention	Intervention	Limitations
Battistone[33] 2016- 1 week USA	Quantitative descriptive	To determine whether an experimental interdisciplinary immersion MSK curriculum would be acceptable, feasible within existing rotations; and effective in strengthening clinical skills.	Residents	MSK Education Week: 5 days which included a Shoulder and Knee Curriculum, Osteoporosis Curriculum, Arthrocentesis/Joint Injection Curriculum, Rheumatology Curriculum and clinical experiences with patients	MSK exam 2	Percentage of trainees reporting ability to evaluate and manage MSK complaints increased (9% to 87% for shoulder; 18% to 86% for knee), and confidence performing MSK injections increased from 10% to 70%. Competency in evaluation of shoulder & knee pain confirmed by OSCEs.	x	Single site, no comparison group, cannot predict proficiency of patient care
Battistone 2016[34] - 3 day USA	Quantitative descriptive	This article describes the first 2 years of the national expansion and implementation of a 3-day MSK Mini-residency.	Physicians, Physician assistants, Nurse practitioner, Nurses in Veterans Affairs (VA) medical centers	3 DAY MSK mini-residency program WITH emphasis on the instruction and assessment of physical examination skills of the shoulder and the knee.	MSK exam 2	The proportion of participants who scored at the 81–100 deciles was 79 % for the shoulder and 65 % for the knee. Mean scores for both stations were high (90 % for shoulder, 86 % for knee, $p < 0.0001$), indicating the success of the program at teaching proficiency in these physical examination techniques.	x	not mentioned
Beran 2012[37] USA	Quantitative descriptive	To assess attitudes regarding teaching of the physical examination in orthopaedic residencies, to assess physical examination knowledge and skills among residents, and to develop a method to track the skill level of residents in order to improve the physical examination curriculum	Orthopedic residents	With the help of subspecialists, we created a checklist of specific PE requirements in which each resident must exhibit competency before completing the rotation. The resident is provided with the appropriate checklist at the start of each rotation. At some time near the completion of the rotation, each attending physician is asked to directly observe the resident performing a complete PE.	MSK exam 4	The overall score of our residents on the OSCE was 66%.	x	Incomplete validation data, low response rate
Dwyer 2013[29] Canada	Quantitative descriptive	To determine if a sports medicine OSCE would demonstrate sufficient reliability and validity to be used for orthopaedic resident in-training assessment.	Orthopedic residents	A compulsory module is a continuous three-month sports medicine rotation, in which residents are expected to acquire the medical knowledge to deal with common sports medicine conditions. Most of the residents undertake two sports medicine rotations, one as a junior resident and one as a senior resident.	Sport Medicine 6	The scores on the total test score were significantly higher for PGY5 residents than for PGY1, PGY2, and PGY3 residents ($p < 0.05$)	x	Bias because - year of some residents known to examiners, low number of candidates, OSCE run over a period of 2 weeks instead of 1 setting, OSCE does not assess surgical skill

Dwyer 2015[30] Canada	Quantitative descriptive	To determine whether junior residents and senior residents could demonstrate clinical skills to a similar level, after a sports medicine rotation	Orthopedic residents	Sports medicine rotation, after which residents are expected to demonstrate competence in the management of patients with acute and chronic soft-tissue injuries of the knee, shoulder, hip, ankle, and elbow. Residents undertake this rotation twice during their training, once as a junior (PGYs 1, 2, and 3) and once as a senior (PGYs 4 and 5).	Sport Medicine 6	There was a significant difference between junior residents and senior residents for each knowledge domain	x	underpowered and bias because year level of some residents will have been known to examiners
Dwyer 2020[32] Canada	Quantitative descriptive	To evaluate a combination of assessment tools used to establish the competence of orthopedic fellows in both clinical skills and surgical performance after completion sports medicine fellowship, with the ultimate goal of creating a certification examination.	Orthopedics fellows	All fellows in a 1-year orthopedic sports medicine fellowship program at the University of Toronto. Fellows undertook 3 4-month rotations with faculty members, with rotations involving a variety of knee- and shoulder-focused practices, as well as hip, ankle and elbow arthroscopy.	Sport Medicine 3 + 4	No significant difference in the mean scores for the checklists or overall GRS between the entry and exit OSCEs.	x	small number of participants, no intraoperative assessment, limited OSCE stations
Dwyer, Slade et al., 2016[31] Canada	Quantitative descriptive	To determine if the use of an Objective Structured Assessment of Technical skill (OSATS), using dry models, would be a valid method of assessing residents' ability to perform sports medicine procedures after training in a competency-based model.	Orthopedic residents	A 3-month sports medicine rotation: at the beginning of the rotation, each resident was provided with a list of 10 technical procedures that they were expected to show competency in at the end of the rotation	Sport Medicine 6	The results show that an OSATS using dry models shows evidence of validity when used to assess performance of technical procedures after a sports medicine rotation. However, junior residents were not able to perform as well as senior residents, suggesting that overall surgical experience is as important as intensive teaching.	x	no intra-operative assessment, no intrarater reliability, no baseline testing of residents was performed
Irwin 2018[36] USA	Quantitative descriptive	To evaluate the long-term impact of an integrated, peer-led, MSUS curriculum for PM&R residents from 2013 to 2017. A secondary goal was to align residents' scores with levels of milestone achievement as the initial step in creating a more	Physical Medicine and Rehabilitation residents	An MSUS curriculum: designed as monthly, peer-led, faculty-supervised, hour-long, didactic presentations that focused on a single joint or body region where image interpretation was linked to pathology and medical decision-making.	MSK ultrasound (MSUS) 2	Most residents achieved the appropriate level of competency for their year	x	no control group, small cohort, no interrater reliability analyses,

		comprehensive, multimodal evaluation that will provide a more accurate view of resident competence.						
Jain 1997[35] USA	Quantitative descriptive	To develop the objectives and to formulate individual cases for an OSCE to objectively evaluate residents' skills in physiatric practice.	Physical Medicine and Rehabilitation residents	Physical medicine and rehabilitation (PM&R) residency	MSK exam 9	Residents' evaluation of the process was very positive. . It was extremely time-consuming and expensive for the faculty to develop the cases.	x	did not assess reliability and validity
McGaghie 1994[38] USA	Quantitative descriptive	To report the interrater and internal consistency reliability of checklist data after their implementation in an OSCE, which is the final examination for a second-year physical diagnosis course.	Medical students -	A second-year physical diagnosis course, before starting clinical clerkships.	MSK exam 4	The 1989 data indicate that, on average, student performance at the OSCE stations ranged from 22% to 44% of the checklist items correct. Student performance improved slightly in the two successive years. In 1990, average performance ranged from 31% to 46% correct . In 1991, the range was from 36% to 49% correct.	x	None reported
Mehrpour 2013[44] Iran	nonrandomized controlled trial	To determine whether (1) a supplemental video educational program enhances performance of medical students' MSK clinical skills and (2) factors such as the proportion of orthopaedic professors to students, sex, age, and previous scores of medical students affected the clinical skills of medical students.	Medical students -	All participants were taught orally and shown how to splint or cast by an orthopaedic surgeon trainer. Both groups had some traditional teaching such as lectures, textbooks, and individual written assignments. In addition one group received the video instructional program.	Splint & cast 1	The medical students who watched the video had an average score of 7.6, whereas the control group's average score was 2.0.	supplemental instructional video	cannot rule out bias from sharing videos or evaluators
Oswald 2011[40] Canada	RCT	To determine if differences exist in MSK PE skills between non-MSK specialist physician and PP®IA taught students physician tutors	Medical students -	Pre-clerkship second year medical in the mandatory Introduction to Internal Medicine (IIM) course where students are taught general PE skills over 12 two-hour small group sessions by a general or sub-specialist physician tutor. An MSK PE small group session replaced one of the 12 usual IIM general PE small group sessions in	MSK exam 5	No significant differences in MSK PE OSCE scores or interpersonal skills OSCE scores between students taught by trained patient educators and those taught by usual non-MSK specialist physician tutors.	teaching by Patient Partners® in Arthritis (PP®IA)	sampling validity, sensitivity of the outcome measures and external validity

				the first week of the IIM block.				
Raj 2006[41] UK	RCT	To compare the core hand and knee examination skills gained by undergraduates taught either by trained patient educators (PEs) or by doctors.	Medical students -	Final year medical students from Nottingham Medical School with no prior formal MSK examination training were randomized for doctor or patient-led teaching followed by a two-station OSCE.	MSK exam 2	There were no significant differences in mean hand or knee OSCE scores.	training from patient educators	Cannot rule out effect of any previous knowledge, cannot prove equivalence with certainty
Siddharthan 2017[45] USA	Case control study	To create and evaluate an OSCE–based MSK workshop designed to simultaneously educate medical students and internal medicine residents, enlisting volunteer medical students as standardized patients (SPs)	Primary Care residents	Yale Primary Care IM residents exposed to a 2-year MSK training curriculum, which consists of a series of lectures linked to interactive workshops scheduled during ambulatory blocks	MSK exam 1	Compared with unexposed residents, residents exposed to the neck/back pain workshop performed significantly better on the clinical skills test	OSCE based MSK (neck/ back pain) workshop	small sample size, limited OSCE stations, outcomes may not translate into improved clinical competence.
Smith 2002[46] USA	Non-randomized controlled trial	To assess student evaluation, satisfaction, and examination outcomes for a new method of teaching MSK medicine clinical skills, structured clinical instruction modules (SCIM), and to compare with the outcomes of a traditional method of teaching clinical skills (small group bedside tutorials).	Medical students -	6-week module on MSK medicine using problem based learning.	MSK exam 2	The SCIM is an effective method of teaching clinical skills in MSK medicine, comparable with patient partners and traditional registrar based bedside teaching methods, but it is less resource intensive.	Structured clinical instruction modules (SCIM)	limited number of MSK stations
Smith 2005[43] USA	Case control study	To create and implement a comprehensive clinical skills teaching model, and to evaluate its effects on residents' knowledge and diagnostic skills	Internal medicine residents	Second- and third year internal medicine residents can elect to join a primary care track, which includes two weekly continuity clinics and a full complement of ambulatory care rotations.	MSK exam 1	Both the shoulder and knee curricula were associated with a significant improvement in test scores, in self-assessment of physical examination, diagnostic and procedural skills, and in OSCE results	curricular series: painful shoulder, and painful knee.	small number of participants, no blinding
Smith 2000[42] Australia	RCT	To evaluate student preferences and examination outcomes of 2 different methods of teaching MSK medicine examination	Medical students - Rheumatology	Students received training in clinical skills in MSK medicine as part of a 4-week clinical skills intensive block in their second year.	MSK exam 1	No statistically significant difference between overall results for medical students taught by rheumatology fellow and by patient partner. No effect of teaching on OSCE result was identified	teaching by Patient Partners	None reported

		techniques: by rheumatology fellows or patient partners						
Chen 2013[48] China	Quantitative descriptive	To describe the use of a pre-validated videotape of an OSCE in a fracture scenario to evaluate raters and to measure the consistency of raters from different subspecialties and with varying levels of seniority	Orthopedics faculty	Orthopedic residency	MSK exam 1	A high Cronbach's α for the 23-item scale regarding global assessment in all 3 tests confirmed construct validity. Kendall's W showed only moderate inter-rater reliability. KR-20 was 0.96 for the pretest, 0.968 for the post-test, and 0.892 for the long-term test, indicating high internal consistency. The p-value for expert validity was 0.626 (independent t-test, n.s.). Using recordings of OSCEs holds substantial promise, it is sufficiently reproducible for use in future training and credentialing purposes.	x	lack of direct interaction between raters and the examinee. Inclusion of raters with varying seniority and clinical subspecialties,
Dwyer, Wright, et al. 2016[49] Canada	Quantitative descriptive	1) To determine credibility and acceptability of modified Angoff method of standard setting, using the Borderline Group and Borderline Regression methods as a reference standard; 2) to determine if it is feasible to set different standards for junior and senior residents, and 3) to determine the desired characteristics of the judges applying the modified Angoff method	Orthopedics residents	Residents undertake the sports medicine rotation twice, once as a junior resident (postgraduate year (PGY) 1–3), and once as a senior resident (PGY 4&5).	Sport Medicine 6	Modified Angoff method is an acceptable method of setting different pass marks for senior and junior residents. This enables both senior and junior residents to sit the same OSCE	x	No alternative standard to establish credibility of method, low number of judges and residents