Assessment of Clinical Competence: Implementation of the Objective Structured Clinical Examination (OSCE) in Saudi Medical Schools

Khaled Mohammad Almisnid

Submitted in accordance with the requirements for the degree of Doctor of Philosophy (PhD)

The University of Leeds

School of Medicine

Leeds Institute of Medical Education

August 2023

Intellectual Property and Publication Statements

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Dedication

To my beloved parents, wife, children, and siblings:

Your boundless love, great sacrifices, and unwavering support have made this remarkable achievement possible. With heartfelt gratitude, this thesis is dedicated to you all.

Acknowledgements

I am profoundly grateful to my supervisors, Dr. Matt Homer, Professor Trudie Roberts, and Professor Richard Fuller. Dr. Homer, my primary supervisor, consistently exhibited an exemplary attitude, offering continuous support and prompt responses to all my academic and administrative needs, making my PhD journey a great experience. Professor Roberts, despite her retirement, remained actively involved in my research, inspiring my thesis with her expansive knowledge and astute insights. Professor Fuller, even post his tenure at the University of Leeds, provided steadfast encouragement, insightful perspectives, and invaluable inputs that significantly enriched my doctoral studies. I am extremely fortunate to have had such an exceptional and supportive supervisory team.

My postgraduate tutors, Dr. Rebecca O'Rourke and Dr. Valerie Farnsworth, deserve heartfelt thanks for their continuous support, enriching workshops, and engaging events that amplified my experience in the PGR office.

I extend deep gratitude towards my home university and the Ministry of Education for their financial sponsorship of my doctorate. I am also thankful to the participants who voluntarily agreed to take part in this project.

Lastly, heartfelt gratitude to my friends Mohammed, Abdulaziz, and my PhD colleagues for their continual support during the highs and lows of this journey. Our shared experiences and memories are truly cherished. Thank you all for your significant roles in my academic journey.

Abstract

Assessment of clinical skills is essential in determining the competency of health professionals. The Objective Structured Clinical Examination (OSCE) is a performance assessment tool that is widely adopted for this purpose. Despite its widespread use in Saudi medical schools, the implementation and impact of the OSCE in Saudi Arabia (SA) have not been widely investigated. The aim of this research is to provide an in-depth understanding of OSCE usage in Saudi medical schools, the opportunities and challenges offered by its adoption, and to develop recommendations to improve implementation.

Using a qualitative constructivist philosophy, this research employs a case study design in two different Saudi medical schools, integrating document review with interviews and focus groups with medical school leaders and their Faculty. Reflexive thematic analysis and codebook methods were applied to interviews/focus groups and documents, respectively.

The analysis generated a series of key themes that impacted the OSCE implementation in Saudi medical schools. These themes constitute a conceptual framework that requires careful consideration in order for them to function harmoniously to produce a high-quality OSCE. This research also highlights that each stage of OSCE implementation in the investigated medical schools involves a series of dilemmas and compromises. Analysis suggests that funding sources (public or private schools), accreditation status, faculty experience, and resource availability all influence the quality of OSCE implementation.

The research findings are consistent with the international OSCE literature; however, this work brings new insights into the use of the OSCE in SA, a non-western culture, and sheds light on the integrative nature of the factors that contribute to successful implementation. It concludes with a series of recommendations that individual medical schools and organisational networks can employ to improve the OSCE in SA, and beyond.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACCLAiM	The Australasian Collaboration for Clinical Assessment in Medicine
AKT	Applied Knowledge Tests
AMEE	Association for Medical Education in Europe
ASME	The Association for the Study of Medical Education
BRM	Borderline Regression Method
DOPS	Direct Observation of Procedural Skills
ETEC	Education and Training Evaluation Commission
EMQ	Extended Matching Questions
FG	Focus Group
HPE	Health Professions Education
ILO	Intended Learning Outcomes
Mini-CEX	Mini-Clinical Evaluation Exercise
MEQ	Modified Essay Questions
MCQ	Multiple-Choice Questions
NCAAA	National Centre for Academic Accreditation and Assessment
NQF	National Qualifications Framework
OSCE	Objective Structured Clinical Examination
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SA	Saudi Arabia
SAQ	Short-Answer Questions
SaudiMEDs	Saudi Medical Education Directives
SCFHS	Saudi Commission for Health Specialties
SDL	Self-Directed Learning
SMLE	Saudi Medical Licensure Examination
SP	Simulated patient
SSME	Saudi Society of Medical Education
USMLE	United States Medical Licensure Examination
WPBA	Workplace-Based Assessment

Presentations and Publications Derived from This Research

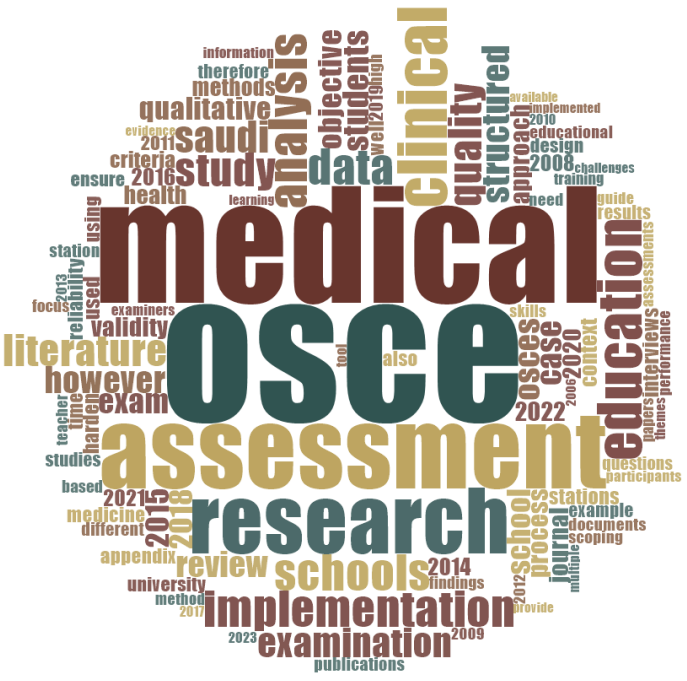
Almisnid, K., Homer, M., Fuller, R. and Roberts T. 2020. Essential Components for Implementing High-Quality OSCEs: A Scoping Review. Oral presentation*. Association for Medical Educators in Europe (AMEE).* The virtual conference.

Almisnid, K. 2020. Refining the Practice of OSCE Implementation: Important Success Factors. Oral presentation. *Clinical Education Network (CLinEdNet).* University of Leeds*,* Leeds, UK.

Almisnid, K., Homer, M., Fuller, R. and Roberts T. 2021. Mastering the Art of OSCEs: Creating Effective Assessment Experiences. Oral presentation. *International Conference on Medical Education (ICME).* Virtual conference.

Almisnid, K., Charnell, A. and Alshobaily, M. 2021. The OSCE in Theory and Practice. Faculty development workshop. *Sulaiman Al Rajhi University, College of Medicine*. Albukairyah, Saudi Arabia.

Almisnid, K., Homer, M., Fuller, R. and Roberts T. 2022. The Necessary Elements for Developing Successful Objective Structured Clinical Examinations. Poster. *The Saudi International Medical Education Conference (SIMEC).* Abha, Saudi Arabia.



The keywords from the data and the thesis — generated by NVivo.

Chapter 1 Introduction

1.1 The rationale of the research

Medical school assessments aim to ensure that medical students acquire the necessary knowledge, skills, and behaviours to practise medicine safely (Dent and Harden, 2013). Assessment is a powerful educational tool that can be used to assess and promote learning (Schuwirth and Van der Vleuten, 2011). Therefore, ensuring the assessment's quality is vital for all stakeholders.

In this thesis, I focus on the OSCE as one of the most common performance assessment tools used in medical schools worldwide (Harden et al., 2015). Over the past four decades, educational assessment literature has highlighted it as a gold standard for performance assessment that offers many advantages and overcomes many drawbacks of other performance assessments, such as long and short clinical examinations (Patricio, 2012; Humphrey-Murto et al., 2013). However, its implementation remains challenging in many parts of the world where it is newly adopted and in contexts where there is a lack of expertise and resources (Ataro et al., 2020; Bearman et al., 2020).

Although the OSCE is widely used in Saudi medical schools (Alsaid and Al-Sheikh, 2017), there is no evidence concerning its implementation or whether it meets the criteria for a 'good' assessment (Norcini et al., 2018; Boursicot et al., 2020). Additionally, there is a rapid proliferation of public and private medical schools in Saudi Arabia (SA) without clear quality control guidelines for assessment (Tekian and Almazrooa, 2011). As a result, the OSCE is implemented differently in different medical schools across the country, with no clear guidelines to guide its use or ensure its quality. Increasingly, Saudi medical educators advocate for quality assurance to ensure that medical graduates' clinical competencies are sufficient to provide safe care to patients (Bajammal et al., 2008; Abu-Zaid et al., 2020).

From my experience as a medical educator and OSCE examiner in various medical schools in SA, I have observed that the OSCE implementation varies considerably between institutions. I asked myself which practices should be regarded as good, what should be improved, and why there is such variation in implementation. However, the Saudi medical education literature provides no answers. This experience has motivated me to undertake in-depth research on this topic to contribute to the field of medical education. Therefore, this project aims to provide empirical evidence about how the OSCE is implemented in Saudi medical schools to develop recommendations that capitalise on the opportunities and overcome the challenges to produce more defensible OSCE outcomes.

1.2 Setting the context for the OSCE

In this introductory chapter, I discuss the need for quality assessment and its criteria. Then, I outline the competencies that medical schools must assess their students on, with a focus on the clinical competency assessment. After that, I will briefly review the OSCE's history and description, focusing on its key benefits and challenges. The second section of this chapter addresses the Saudi context and provides background information on medical education in SA and how it is regulated before concluding with a review of the status of undergraduate and postgraduate clinical assessment in SA.

1.2.1 The need for quality assessment

Assessment in education refers to the process of testing, evaluating, measuring, or collecting data or evidence about someone's knowledge or performance (Norcini et al., 2018). It is considered the tool that enables educators and institutions to make decisions on student progress and judge institutional outcomes. Assessment can perform multiple functions, such as encouraging learners to learn (Miller, 1990), measuring students' competencies, informing educators about students' performance, and informing institutions about their outcomes (Case and Swanson, 2001; Colliver, 2002; Cottrell, 2006; Norcini et al., 2018). Effective assessment practices not only drive students to learn and enable assessors to measure that learning, but they also draw students' attention to the core knowledge that they need to master (van der Vleuten et al., 2010). Therefore, assessment systems should be carefully designed (Wass et al., 2001).

Furthermore, good assessment should produce defensible decisions, for it has paramount importance and significant consequences for all stakeholders (examinees, examiners, educational institutions, regulatory bodies, and the community) (Downing, 2003; Amin et al., 2006). Thus, assessment should be designed according to evidence-based practice and tested scientific criteria and should employ credible methods to produce defensible decisions.

1.2.2 Criteria of quality assessment

Certain criteria are used to ensure the quality of assessment techniques (i.e., a high-quality assessment that produces defensible decisions). The need for defensible decisions is underpinned by the need for a clear explanation of how results are generated and how they can be interpreted (Kane, 2010). However, no single set of criteria can work with all assessment types and purposes (Norcini et al., 2018). Thus, it is important to look for evidence (such as policies, documents, and procedures) on which to base assessments, use different assessment tools, and determine which criteria need to be fulfilled according to the intended purpose of that assessment to correctly design and

implement a good assessment (Cook et al., 2014; Boulet and McKinley, 2022; Boursicot et al., 2022).

Nonetheless, there are some general principles to consider in most assessments. van der Vleuten (1996), Amin et al. (2006), and Schuwirth and van der Vleuten (2014) specify five criteria that quality academic assessments should meet: assessment should be valid, reliable, feasible, fair, and useful for the learning process. Consensus statements and recommendations from the Ottawa 2010 Conference identify some important assessment criteria as validity, reliability, educational impact, catalytic impact, feasibility, fairness, and acceptability (Norcini et al., 2011; Norcini et al., 2018; Norcini and Ben-David, 2021). Table 1 shows the most important assessment criteria and describe their meanings. Effective assessment methods have several facets; neglecting any of these facets can decrease assessment quality and attenuate its decision defensibility.

Table 1. Criteria of assessment and their meaning.

Criteria	Meaning	Types
Validity or coherence	The assessment measures what it is supposed to measure and reflects the curriculum.	Propositions, evidence, arguments, and decisions' validity (Messick, 1989; Kane, 1992) OR classical types: content, construct, face, and criterion validity (which are two subtypes: predictive and correlational "concurrent" validity).
Reliability, reproducibility, or consistency	When the same assessment is introduced again (in similar circumstances), it should produce similar results.	Inter-rater, internal consistency, inter-station, and test-retest reliability.
Feasibility or practicality	The practicality of assessment in terms of resources required for implementation	Feasibility in terms of time, cost, efforts, logistics, workforce, and examinees.
Fairness or equivalence	The assessment results lead to similar or equivalent decisions across equivalent cohorts or institutions.	Procedural and substantive fairness.
Educational impact or effect	The educational benefit that students gain from taking the assessment.	
Catalytic impact or effect	The motivation that all stakeholders gain from assessment results, and feedback that supports and enhances future education.	
Acceptability	That different stakeholders recognise the assessment's importance and credibility.	

(van der Vleuten and Schuwirth, 2005; Amin et al., 2006; Kane, 2010; Norcini et al., 2011; Cook, 2014; Cook et al., 2014; Harden et al., 2015; Norcini et al., 2018; Dent et al., 2021; Norcini and Ben-David, 2021).

Some differences exist among researchers in their assessment interpretations and approaches to any of the aforementioned criteria. For example, some authors define validity broadly to include fairness (Kane, 2006), and others define fairness broadly to include validity (Kunnan, 2000). However, Kane's (2010) view is that validity and fairness are different concepts, but the overlapping similarities are more prominent than differences; this interpretation seems to me more plausible. Although there is no consensus on any definition for any assessment criteria, there is a generally accepted understanding of the concept for each of them that assists assessors in using them effectively.

Utilising these criteria correctly (along with other psychometric tools) can help assessors produce defensible results and increase their confidence in their decisions. For instance, procedural fairness or relevant standardisation is an important aspect of assessment and ensures that similar examinees should take the same test under the same conditions, be treated in the same manner, and have their assessment evaluated and interpreted in the same way (Kane, 2010). Keeping this in mind when assessment tools are prepared, implemented, and evaluated helps to ensure the fairness of the assessment. For example, if students are evaluated using an OSCE for their history-taking skill, and some examine a well-trained simulated patient (SP), and others examine an untrained patient, the outcomes may be unfair. Thus, standardised assessment processes should produce more trustworthy results.

The process of designing effective assessments that fulfil the aforementioned criteria necessarily includes multiple phases. It starts with blueprinting because while no single assessment can evaluate all curriculum, competencies, or intended learning outcomes (ILOs), an effective assessment must produce a valid, manageable and feasible exam that sufficiently examines the competencies of the course as a whole (Downing, 2003; van der Vleuten et al., 2010). This step ensures a match between curriculum and assessment and avoids unintended mismatching (Newble, 1998; Trigwell et al., 2000). Therefore, blueprinting is central to supporting authentic content and assessment validity.

The next phase involves determining the form of assessment, choosing the appropriate tool for the purpose, and designing test items (Downing, 2006). This step is lengthy, consisting of several smaller steps, such as identifying the issue to be examined, constructing model answers, revising with experts, proofreading, and trialling the assessment beforehand (Jolly, 2014). The next stage is to administer the assessment in a controlled environment by eliminating (or at least minimising) sources of error and then

combining all scores from different types of assessments using empirical evidence and logical argument to produce a global score (Downing, 2003).

The process of developing a quality assessment is not finished after administering the exam and then releasing the results. The subsequent stage is the post-exam analysis, which includes qualitative (Tavakol and Dennick, 2011b) and quantitative (psychometric) aspects (Pell et al., 2010), both of which are important to ensure the validity of the internal structure of exam questions. It is also useful in identifying certain flows in assessment, developing later assessments, and enabling assessors to provide individualised feedback to examinees and assessment writers (Lane et al., 2016).

These phases may appear to be challenging for real practice as they are time- and resource-consuming, and not all assessments require all steps to be performed; however, each step can play an important role in contributing to the validity of the assessment. Furthermore, it is better to conceptualise designing an assessment as an iterative rather than a linear process. Medical schools need to be systematic and inform their assessment designs through evidence-based practice guidance to make correct decisions and produce defensible and trustworthy results. This will improve medical graduates' clinical skills and better prepare them for future practice.

1.2.3 Assessment of medical competencies

Medical schools strive to ensure that their graduates have the competencies necessary to safely deal with patients and meet health sector requirements. In order to achieve this goal, educators usually categorise these competencies into the three domains of knowledge, skills, and attitudes (Harden et al., 2015). Medical educators wish to assess these competencies as best as possible to ensure that they graduate competent and safe doctors. As a result, the field of assessment has evolved, as old assessment methods have been improved or replaced by new assessment tools to meet the high-stakes nature of medical assessment (Harden and Gleeson, 1979).

To assess knowledge, applied knowledge tests (AKTs) (written assessments) are used, which only assess the first two levels of Miller's (1990) pyramid of 'knows' and 'knows how' (Figure 1). For example, Multiple-Choice Questions (MCQs), Extended Matching Questions (EMQs), Short-Answer Questions (SAQs), and Modified Essay Questions (MEQs) can assess students' knowledge and understanding (Wass et al., 2001). Medical educators and psychometricians are reasonably satisfied with advances in this field because the assessment of knowledge has received attention for a reasonably long time and from all disciplines, supporting those advances (van der Vleuten and Schuwirth, 2005).

When educators try to assess attitudes and behaviour, however, they face challenges, finding these qualities more difficult and intractable to assess. Many assessment

researchers think that attitudes such as professional demeanour are difficult to measure psychometrically, but others argue they can be evaluated using workplace-based assessment tools (WPBA) such as 360-degree assessment (Whitehouse et al., 2007; Hodges et al., 2011; Hodges et al., 2019).

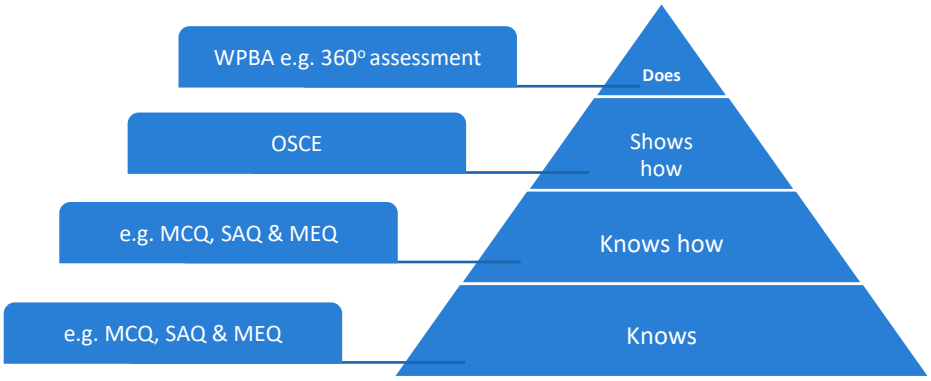


Figure 1. OSCE in relation to Millar's pyramid of assessment.

The third student competency that medical educators need to assess is their clinical communication and performance skills. Assessing clinical competencies is quite complicated, making these skills more difficult to assess than knowledge but perhaps easier to assess than attitudes (Khan et al., 2013b). Moreover, ensuring that medical students obtain minimum competency in clinical skills is crucially important for medical schools, regulatory bodies, and the health profession in general because of the high-stakes nature of the medical field and the serious consequences for all parties if the competencies are not achieved. To assess students' skills and clinical performances, which rests higher in Miller's pyramid at 'shows how' level, requires performance assessment tools (van der Vleuten and Schuwirth, 2005; van der Vleuten et al., 2010; Etheridge and Boursicot, 2021), such as OSCEs, Mini-Clinical Evaluation Exercise (Mini-CEX), and direct observation of procedural skills (DOPS). Development of the tools that assess these skills is perhaps not as advanced as that of tools that assess AKT because of the complexity of the skill area and the relative newness of this type of assessment compared to AKTs (Epstein, 2007; Boursicot et al., 2011; Boursicot et al., 2020).

In short, medical schools are looking to ensure that their students adequately possess the three important competency components of knowledge, attitude, and skills. Knowledge is relatively easy to measure, but attitude and behavioural measurement are challenging, entangled and difficult to measure. Skill measurement is also difficult but still possible and manageable. Competent clinical skills are crucially important for doctor

and patient safety. Therefore, the assessment of clinical skills requires an accurate and highly evolved assessment tool.

1.2.4 Assessment of clinical skills

In the past, clinical competencies were assessed by long and short clinical examinations (Harden et al., 2015). This traditional clinical examination involved asking medical students to examine one long clinical case (clinical viva) for about an hour (van der Vleuten, 1996), usually unobserved (Ponnamperuma et al., 2009). Then, the candidate was assessed by one or two medical doctors for typically about half an hour (Gleeson, 1997). This method served as the dominant clinical assessment tool and has some advantages; for example, it is authentic and provides a holistic approach to patient care (Wass and van der Vleuten, 2004). However, it has many obvious limitations as well (Harden et al., 1975). For example, the sampling and case-specificity of content addressed in the assessment are key for the validity and reliability of educational assessment (van der Vleuten, 1996). Indeed, the medical student may be well-versed in the content related to his/her assigned clinical case, so the student will get a high grade on this assessment, but he or she may know little or even nothing about other cases (Ponnamperuma et al., 2009). The opposite is also true, i.e., the student may encounter a case that he or she knows nothing about and therefore fail the exam, even though this student may master most of the other material. The assessor assumption in these scenarios is that the student in the first example will do well in real-life situations with real patients, and the student in the second scenario will not. These assumptions are not necessarily true or fair, as biases and sampling problems would directly attenuate assessment reliability and credibility (Ponnamperuma et al., 2009; Vincent et al., 2022).

Another limitation of this method is that relying on the judgment of one or even two assessors does not support a fair process due to the inherent bias that all assessors have, impacting the validity of any given decision (Wass and van der Vleuten, 2004). Also, this assessment is poorly structured, with no clear scoring method (Khan et al., 2013b). Similarly, short case examinations, during which usually a small number of cases are assessed by one examiner, struggle largely with the same limitations that exist for one long clinical examination. Short- and long-case clinical examinations lack standardisation, structure, and sufficient sampling (Khan et al., 2013b)

1.2.5 The introduction of the OSCE

In 1975, Harden et al. introduced the notion of the objective structured clinical examination (OSCE) to overcome some limitations of traditional assessment; this concept was further developed by Harden and Gleeson in 1979. The guiding principle behind the OSCE was to have more samples for content to be examined and allow candidates to perform more skills in multiple stations (Harden, 1988; Harden, 1990;

Harden et al., 2015). Furthermore, because all assessors have inherent biases, candidates are seen by a number of different assessors, which can diminish the bias effect (van der Vleuten et al., 2010; Gingerich et al., 2014). Hence, institutions can be more confident about the validity of the candidate's global grade and future performance. For example, suppose the students went through 18 different stations, and all 18 assessors rated the student as very good. In that case, the school can be more confident in predicting that this student will perform at that same level on a different day with a different patient and vice versa. Figure 2 depicts the difference in sampling and other main differences between the OSCE and the long-case clinical examination.

The OSCE method of assessment attempts to overcome some of the limitations of traditional clinical assessments by improving the validity and reliability of the exam (Khan et al., 2013b; Abdulghani et al., 2015). It provides better sampling for clinical examinations rather than one case. It collects grades or opinions of multiple assessors, instead of one or two, to provide more valid, reliable, and fair judgment on students' performance with less room for bias.

Due to the fact that few practical tools exist to assess medical students' clinical skills, as well as the OSCE's benefits and adaptability, it has been widely adopted by medical schools around the world (Norman, 2002; Humphrey-Murto et al., 2013; Khan et al., 2013b; Daniels and Pugh, 2018). Many authors, including Khan et al. (2013b), Harden et al. (2015), and Boursicot et al. (2020), believe that a well-structured OSCE is one of the best tools to assess clinical competencies. However, OSCEs, like all assessment tools, need to comply with the criteria of good assessment (as discussed in Section 1.2.2) in their design to produce high-quality results (Patricio, 2012; Zabar et al., 2013). Those criteria should be carefully considered, as assessment in medical schools is regarded as a high-stakes undertaking due to its significant impact and influence on students' futures, communities, and direct patient safety. For instance, a non-competent student who successfully passes a final year clinical exam—and who should have failed—will have an unfavourable impact on society and patient safety. Thus, assessment in medical settings should be designed and carried out in a fashion that allows only competent students to pass (Norcini et al., 2011).

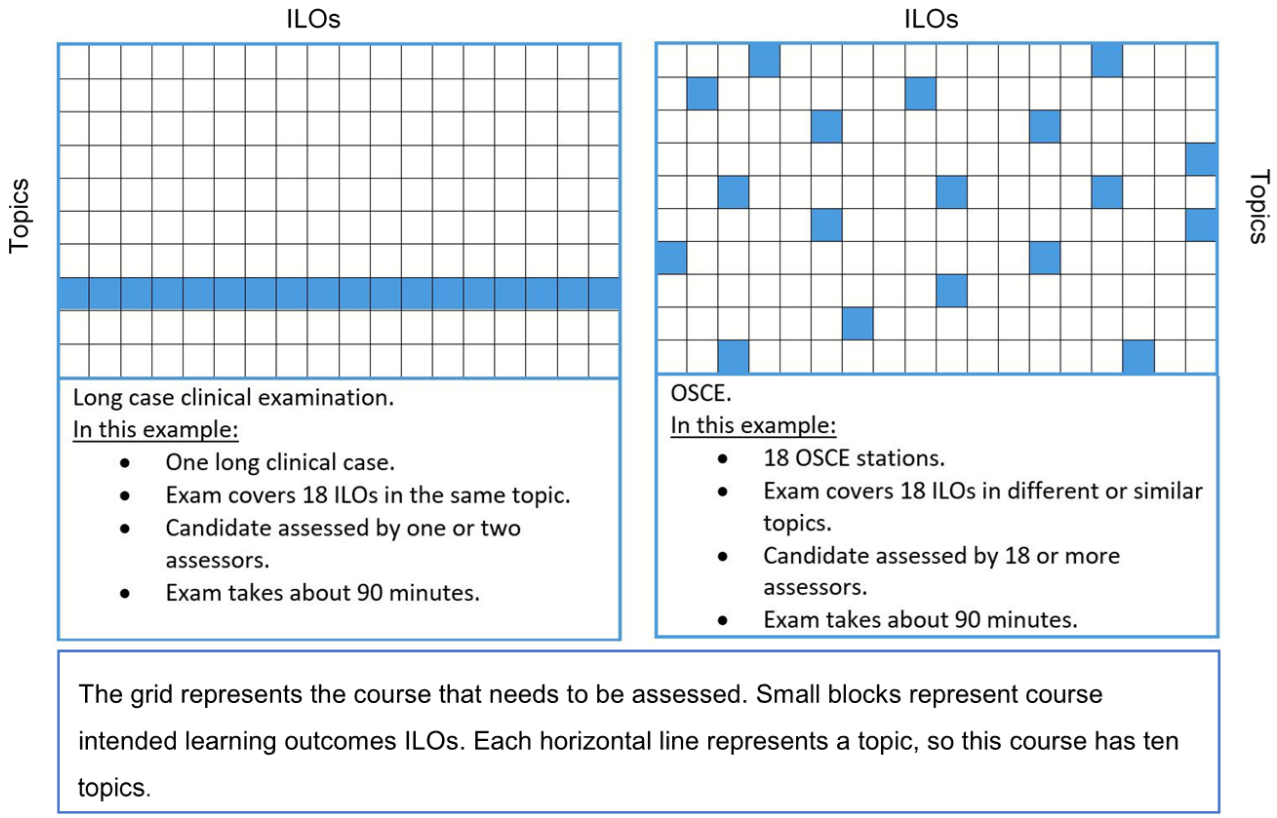


Figure 2. The key differences between the long case clinical examination and the OSCE.

1.2.6 Description of the OSCE

Harden (1988) describes the OSCE as a performance-based examination designed to measure students' clinical skills and competencies in an objective and structured manner. OSCEs are objective assessments, as each candidate goes through the same set of stations (Harden et al., 1975; Harden and Gleeson, 1979). It is structured so that each assessor's score follows a unified marking scheme, which could be a checklist or a rating scale (Gupta et al., 2010). Assessors have to observe candidates' performances and complete the marking scheme; these schemes may include a global score and assessors' written feedback based on the performance of each candidate after every station (Boursicot and Roberts, 2005). OSCEs are also structured so that they test a specific set of skills in each station. It can cover the four main areas of history taking, clinical examination, procedures, and explanations (Harden et al., 2015). Hence, it typically requires a student to demonstrate competencies by performing a skill rather than by describing a theory.

In OSCEs, students go through multiple stations in a circuit fashion; each station typically takes between five to twelve minutes to complete (Boursicot and Roberts, 2005). Students are prompted to move on to the next station by some form of a signal. The circuit may also include rest stations (Figure 3). The number of stations and duration in

each station is largely dependent upon the content blueprint and overall exam time (Khan et al., 2013a). In each station, the candidate encounters a question or instruction, e.g., measuring blood pressure. Each station includes a real patient, simulated patient (SP) (actor), or manikin.

OSCEs can be conducted in a hospital setting, e.g., for final-year students or in a special venue professionally prepared to mimic real-life experiences. The station's layout may vary according to the location and skills to be performed. Any equipment required should be prepared beforehand.

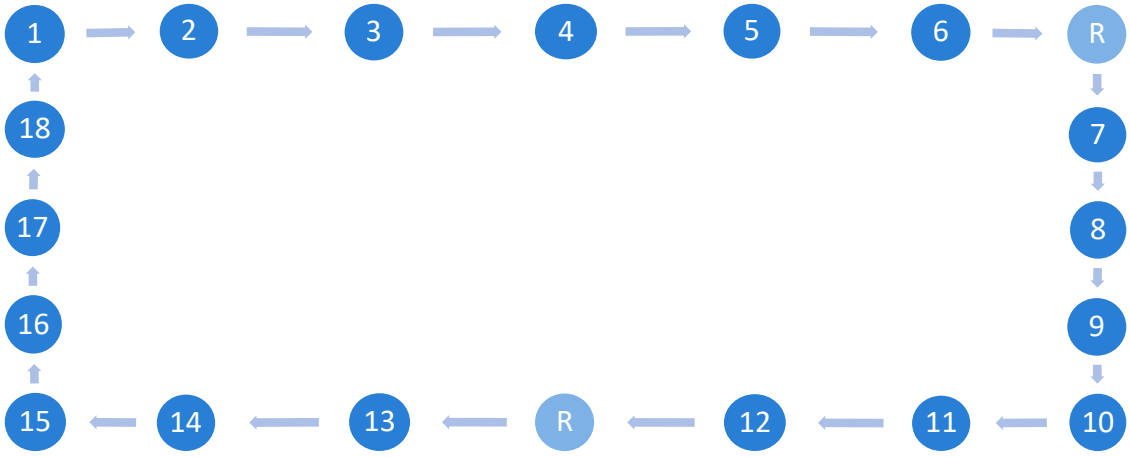


Figure 3. OSCE stations layout in the exam venue.
Examiners stay in the station, but candidate moves around. Numbers indicate station order and (R) indicates a rest station.

1.2.7 Comparing the OSCE to good assessment criteria

OSCEs are flexible assessment tools, i.e., they can accept modifications to enable continuous improvement (Harden et al., 2015). However, the literature does not address the fact that the quality and delivery of OSCE are often dependent on the experiences of its designers and examiners (Daniels and Pugh, 2018). In the following sections, I present and discuss some advantages and disadvantages of the OSCE.

1.2.7.1 Advantages of using the OSCE as an assessment tool

If OSCEs are designed and implemented properly, they should have a high level of validity (Downing, 2003) and positive educational benefits (Jafarzadeh, 2009; Boursicot, 2010; Harden et al., 2015). For the sake of this high validity, the OSCE designer should be a content expert in a given speciality (Selby et al., 1995). The OSCE designer has to blueprint the ILOs of the targeted curriculum and ensure that a variety of skill domains is assessed (Boursicot and Roberts, 2005). Blueprinting could be done using a two-dimensional matrix with one axis focused on the competency to be tested, e.g., history taking, and the other axis focused on the course or block that needs to be addressed,

e.g., respiratory system (Table 2) (Gupta et al., 2010). The number and topics of stations in an effective OSCE should be determined and sampled proportionally against the whole intended content to examine. Another source of evidence for OSCE validity is that its scores are highly correlated with other related assessments (Harden et al., 1975; Simon et al., 2007; Orovec et al., 2022).

Table 2. An example of an OSCE blueprint matrix.

Domain / Course	History taking (Information gathering)	Examination	Procedure	Explanation (Information giving)
Gastrointestinal	Abdominal pain	Liver palpation	_	Read and explain liver enzymes results
Cardiovascular	_	Heart auscultation	Take blood pressure	Read and explain an ECG
Respiratory	Asthma	Chest examination	Introduce a nasal tube	Read and explain chest X-ray
Musculoskeletal	Headache	Knee examination	_	_
Neuro	_	Neurological examination of legs	Otoscope	_
Reproductive	Amenorrhea	Breast examination	_	_
Generic	_	_	Inserting cannulation	Breaking bad news
Total number of stations/questions	4	6	4	4

Another advantage to OSCEs is that the designer can control their difficulty and complexity (Harden and Gleeson, 1979). The designer can utilise SPs and train them on how to act or respond to students during the exam (Daniels and Pugh, 2018). For example, the SP can be instructed to play a straightforward case for junior students, a difficult case for advanced-level students, including content that may focus on sensitive or challenging cases or highlight areas of cultural importance (Altshuler and Kachur, 2001; Parish et al., 2006).

One of the basic tenets of the OSCE is that each candidate is assessed in multiple stations by multiple assessors. All candidates should be assessed by the same set of stations and, ideally, by the same assessors. Large numbers of stations and different assessors in each station enhance results reliability (Roberts et al., 2006; Brits et al., 2021). When a candidate is evaluated by a large number of assessors, variation between assessors should be kept at a minimum to enhance reliability (Harden and Gleeson, 1979; Gormley, 2011). van der Vleuten (1996) empirically demonstrated that bias severely affects reliability in the case of only one examiner assessing the examinee in one or even multiple clinical cases. He also demonstrated that using multiple stations and a new assessor for each station will dramatically improve assessment reliability and

fairness and, ultimately, assessment acceptability. OSCEs are far better in maintaining these criteria than the traditional long case clinical examination.

The use of a carefully constructed marking scheme would support the validity of the OSCE by aligning it with the ILOs and specifying the areas that the assessor has to inspect during candidate performance (Abdulghani et al., 2015; Donohoe et al., 2020). This structured and clear scoring rubric can minimise prejudice and variation between examiners and enhance reliability (Khan et al., 2013b; Harden et al., 2015).

The OSCE is considered fairer than traditional clinical examinations in terms of standardisation, validity, and reliability (Harden et al., 1975; Khan et al., 2013a; Cho et al., 2022) because the three factors (SP, examiner, and case) can be highly controlled in the OSCE. That means OSCE can be designed to standardise the station case, the SP, and the examiner in each station so that, in theory at least, the only source of variation in scores is due to the candidate, ensuring that the assessment produces highly reliable scores (Figure 4) (Harden and Gleeson, 1979).

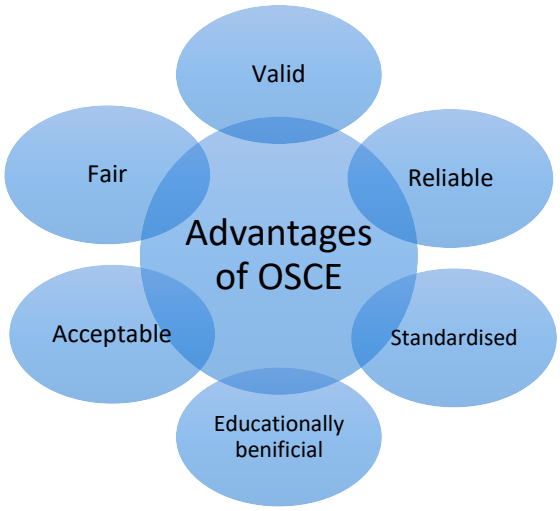


Figure 4. Summary of main OSCE advantages.

1.2.7.2 Challenges of using the OSCE as an assessment tool

In terms of cost, time, place, patients, and examinee/examiner ratios, OSCEs can seem unfeasible and resource-intensive (Carraccio and Englander, 2000; Barman, 2005; Rushforth, 2007; Patricio et al., 2013; Bearman et al., 2020; Zamanzadeh et al., 2021). This problem is more prominent in summative end-of-year assessments because of the need for adequate numbers of stations, SPs, assessors, and logistical support personnel, such as station organisers, timekeepers, and invigilators. There is also a need for sufficient space, and the exam is lengthy, so it may also be costly (Jiang et al., 2021).

These factors make OSCEs difficult to design and consume much time and effort to arrange (Cusimano et al., 1994; Abdulghani et al., 2015). Harden et al. (2015) indicate that most of this complexity and time will be spent before commencing the exam, so when the exam starts, the examiners' time will be used more efficiently. However, it is one of the best available high-stakes clinical assessments, so its overall value should outweigh its limitations (Patricio et al., 2013). Moreover, assessing higher-level thinking skills (e.g., application of skills) would need more than a simple tool that measures straightforward knowledge. The nature of this domain (clinical skills) requires a sophisticated assessment instrument to accurately measure skill competencies (Walsh et al., 2009).

Another challenge for the OSCE is the examiners' variation in stringency, especially in OSCEs with parallel circuits, which reduces the OSCE's reliability and fairness (Boursicot et al., 2020; Yeates et al., 2021; Homer, 2022). Multiple factors can influence the examiner's judgment and result in unfair marking. These factors include the examiner's cognition (Gauthier et al., 2016); the contrast effect between candidates (examiner judgment on one candidate's performance influenced by previous candidate performance) (Yeates et al., 2015b); the nature of the examiners '*hawks'* and '*doves'* (Bartman et al., 2013); and the examiner's experience in assessing OSCE candidates (Chong et al., 2017). However, unwanted examiner variation can be reduced by providing adequate training and sampling (Pell et al., 2008; Harden et al., 2015; Fuller et al., 2017; Yeates et al., 2019)

OSCEs have been criticised for assessing only isolated skills and compartmentalising approaches to cases rather than taking a holistic approach as is needed when caring for real patients (Harden and Gleeson, 1979; Gleeson, 1997; Gupta et al., 2010). This indicates a need to combine other clinical assessment types besides OSCEs to work more in an integrated way (Khan et al., 2013b). No single assessment tool would be able to assess all aspects of any curriculum or clinical aspect (van der Vleuten, 1996), and OSCEs are not an exception. Therefore, OSCEs should be utilised alongside a range of different assessment formats that integrate to produce the best possible outcome, for example, by assessing the 'Does' level of Miller's pyramid and have more authentic assessments (Ben-David et al., 2001; van der Vleuten et al., 2010).

In brief, OSCEs are flexible assessment tools that incorporate effective assessment criteria (Turner and Dankoski, 2008). This can be done in a variety of ways. Careful consideration for adequate sampling, station design, and professional implementation can produce high-quality OSCE and defensible results. On the other hand, no accurate blueprinting and bad sampling often produces unreliable scores, resulting in resource waste and inaccurate decisions about candidate competence.

1.3 Saudi context

This section provides background on medical education in SA and the role of education regulatory bodies. Then, I briefly discuss the contemporary clinical assessment status in under- and postgraduate education in SA.

1.3.1 Medical education in SA

The first Saudi state was founded in 1727 (MOFA, 2023), and the first record of Western (modern) medicine being practised in SA was in 1926 when the Health Directorate of Makkah was established during King Abdulaziz Al-Saud's visit to the Hejaz region (Mufti, 2000). Between 1727 and 1926, the health practice in SA relied on traditional medicine and a few doctors from outside the Arabian Peninsula (Wahba, 1935; Attia, 2022). From 1926 onwards, there was increased interest in modern medicine as doctors from neighbouring countries were drawn to work in SA and later by government scholarships for Saudis to study medicine abroad (Telmesani et al., 2011; Almasabi, 2013). The Saudi government decided to improve the health system by establishing Saudi medical schools and graduating Saudi doctors, so between 1967 and 1996, five medical schools were established (Al-Rabeeah, 2003; Zaini, 2007).

Although medical education is a relatively new discipline in SA, as the oldest medical school in the country was established in 1967, the pace of development in this discipline has been highlighted as very promising (Bin-Abdulrahman, 2011). Until 2000, there were only five medical schools in the country; by early 2023, that number had climbed to 41 (Alrebish, 2014; Alamro et al., 2023). Figure 5 shows the distribution of medical schools in SA (GaStat, 2022; NCAAA, 2023b). This rapid growth in the number of medical schools has raised concerns about the quality of medical school graduates (Smith and Abouammoh, 2013; Abu-Zaid et al., 2020). There is an awareness of the importance of ensuring quality as well as quantity to ensure that graduates are competent doctors (Telmesani et al., 2011). Medical school graduates must be able to contribute to the ambitious Saudi Vision 2030 regarding the health sector in the country by providing health services at a very high level that matches the levels reached in developed countries (Vision2030, 2023). One of the goals of the Saudi vision is to privatise some of the public medical schools and enable them to develop a financially sustainable operational system and create a competitive environment to improve their outcomes. To achieve these ambitions, a core activity of all medical schools is the design and delivery of rigorous, high-quality structured assessments to ensure they are graduating highly qualified doctors who are competent enough to enhance and develop health services to the desired level.

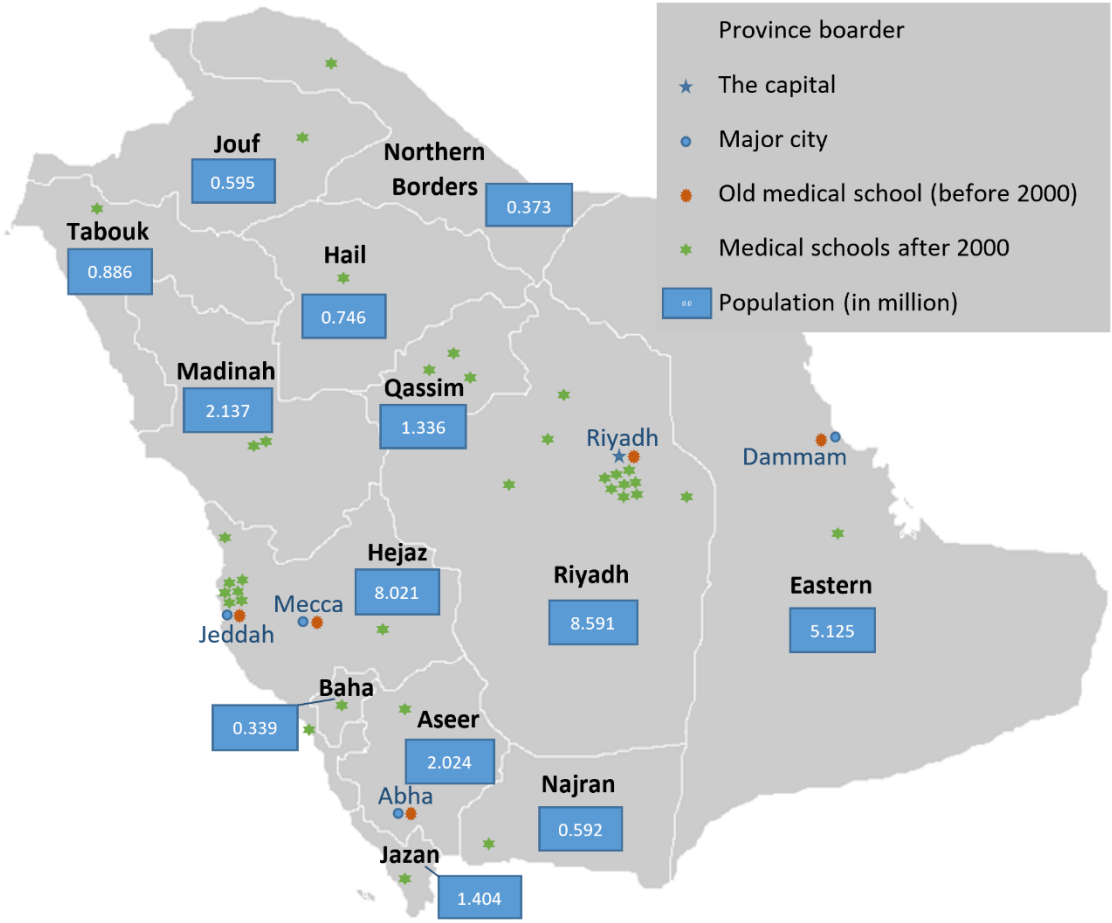


Figure 5. Distribution of medical schools in SA as of 2022.

The awareness of medical education importance is growing markedly, and the need for more medical educationists is becoming essential to match the growth of medical schools (Bin-Abdulrahman et al., 2012). Thus, medical schools now pay more attention to medical education, and most of them have established a Department of Medical Education that regulates the internal educational process to ensure the programmes remain current in cutting-edge research and practises (Al Shawwa, 2012). There is a growing call to prioritise medical education research in SA, with a focus on culturally relevant research (Abdulrahman, 2012; Obeidat et al., 2015). Medical educationalists within their departments are expected to develop their medical schools, support teaching, learning and research, and continuously assess and develop medical students' competencies in terms of knowledge, skills, and professional behaviours (Tekian and Almazrooa, 2011). Medical schools in SA apply different teaching approaches, such as traditional, problem-based learning (PBL), team-based learning (TBL), and hybrid systems that combine two or more systems (Ibrahim et al., 2014). In general, there is a move towards student-centred education and self-directed learning (SDL) (Alghasham, 2012).

1.3.2 Regulatory bodies in SA

National education regulations in SA are governed by the Education and Training Evaluation Commission (ETEC), which has two bodies to regulate higher education in SA (ETEC, 2023a). The first is the National Qualifications Framework (NQF), which is a framework that aims to develop and unify the quality of educational institutions in SA to produce educational outcomes that meet national and international recognition (NQF, 2020). The second, the National Centre for Academic Accreditation and Assessment (NCAAA), ensures that higher education institutions (such as medical schools) meet NQF requirements by encouraging improvement, evaluating quality, providing accreditation, and promoting international recognition for programmes and graduates (NCAAA, 2023a). The NCAAA uses the Saudi Medical Education Directives Framework (SaudiMEDs) as a framework for the competencies that graduates must attain (Alrehaily et al., 2022). The SaudiMEDs framework outlines the minimum competencies that all Saudi medical graduates must possess in order to practise safely.

One of the core responsibilities of these two entities (NQF and NCAAA) is to ensure that assessment in higher education institutions is valid and reliable according to their criteria. For example, they look for the integration of appropriate assessment methods into the learning experience in the appropriate domain of learning outcomes to produce high-quality assessments (NCAAA, 2023a). Accredited medical schools need to prepare an annual progress report to follow up on any significant changes to their programme and submit it to the NCAAA (ETEC, 2023b). The NCAAA will not visit the school for at least five years from the date of accreditation.

Medical education departments within each medical school focus on improving instruction, curriculum, assessment, and other related issues according to international development in the field (Bin-Abdulrahman et al., 2012; Cleland and Roberts, 2021), within the broader limits of NQF, NCAAA, and Saudi Medical Education Directives Framework (SaudiMEDs framework) requirements (Bin-Abdulrahman, 2011). Other bodies, such as the Saudi Medical Colleges Deans' Committee and the Saudi Society of Medical Education (SSME), monitor, participate in, and support medical education development in SA (Bin-Abdulrahman, 2011; SSME, 2023).

1.3.3 Undergraduate clinical assessment in SA

The assessment approach applied in any educational institution has an extensive influence on students' approach to learning (Entwistle and Ramsden, 2015). It has been obvious, at least in the context of Saudi culture and student educational background, that the assessment type, i.e., summative assessment type, strongly influences their approach to learning. I observed this first-hand in my own studying and subsequent teaching experience in medical schools; for example, I noticed that students use "exam-

oriented reviewing" or "exam-oriented reading" terms to advise each other on strategies for studying and learning. These terms reflect the influence of assessment on their learning process and that their primary focus is often on exam performance, not on continuous learning or gaining and retaining knowledge. This is why carefully choosing an assessment system is important and why continuous evaluation and improvements are crucial to the educational process (Boud, 2000; Case and Swanson, 2001; van der Vleuten et al., 2010; Tekian and Almazrooa, 2011; Boud and Soler, 2016).

In general, summative assessment is dominant in Saudi medical schools, while little attention has been given to formative assessment (Amin et al., 2011; Alrebish, 2014; Al-Wassia et al., 2015). MCQs, EMQs, SAQs, and modified essay questions (MEQs) are the most common assessment tools in Applied Knowledge Tests (AKT) exams. For clinical assessment, according to some authors, many schools overly rely on one or two assessment tools, mainly objective structured clinical examinations (OSCEs), and occasionally the mini-clinical evaluation exercise (MiniCEX) (Amin et al., 2011). Some medical teachers believe that they do not utilise or apply OSCE correctly; the evidence suggests that medical teachers do not question the tool itself, but they do question its application (Bakhsh et al., 2009; Alghamdi et al., 2016; Hadi et al., 2018).

1.3.4 Licensure and postgraduate clinical assessment in SA

The Saudi Commission for Health Specialties (SCFHS) conducts the Saudi Medical Licensure Examination (SMLE), which is a 300-item MCQ-based substantial AKT exam that places an extra burden on students, even before they graduate, as it is mandatory for all who want to practice, specialise in, or obtain a clinical position in SA (SCFHS, 2023). However, many SA medical educators advocate for adding another component that examines the clinical competencies of new graduates (Bajammal et al., 2008; Abu-Zaid et al., 2020). While this would add a layer of complexity for newly graduated students, as they will be required to pass both parts of the SMLE to obtain a medical licence, it would ensure the competency of newly graduated doctors and reassure the community and other stakeholders. Thus, preparing medical students for the potential clinical part of the exam has become necessary, making the need for a more valid and reliable assessment at medical schools more important now than ever before.

Medical schools often support their students in taking the AKT part of this exam. For example, some medical schools offer lectures on how to prepare for SMLE and organise and administer formative AKT exams mimicking the real SMLE exam in the type of questions and duration. However, this is not yet the case with the proposed clinical exam. As with the first part of SMLE, medical schools need to prepare their students for this new assessment to ensure that their students are well-prepared for postgraduate high-stack exams. Moreover, the problem of clinical assessment seems more complicated than the AKT part of the exam due to the complexities of assessing clinical and

interpersonal skills, as well as the lack of a national guideline that supports clinical assessments. (Bakhsh et al., 2009; Abu-Zaid et al., 2020). Therefore, medical schools must improve their own practises for implementing high-quality OSCEs to prepare their graduates with the knowledge and skills required to perform well on the current and potential SMLE parts and graduate better doctors.

From my observation, in addition to the official use of the SMLE, it is also unofficially used as a benchmark for evaluating the outcomes of medical schools and as a determining factor for establishing a school's reputation for the quality of teaching and assessment processes. Thus, this licensure examination is important for all stakeholders, including applicants, students, medical teachers, decision-makers, and the community in SA. Some Saudi medical graduates tend to continue their education or training outside the country. Then, they need to pass the host country qualifications, such as the United Kingdom (UK) Professional and Linguistic Assessments Board test (PLAB), which contains a clinical component (GMC, 2023b). Thus, medical schools need to better prepare their graduates for national and international licensure exams.

1.4 Summary

In this chapter, I argued the necessity of quality assessment in medical education and analysed the clinical assessment challenges that existed prior to the development of the OSCE as well as the OSCE's conceivable benefits. I reviewed the criteria for good assessment and demonstrated how the OSCE meets them by describing its process, advantages, and disadvantages. Then, I established the context of medical education in SA and emphasised the need for quality assessment to ensure the quality of its outcomes in light of the recent and rapid growth of medical schools in the country. In the following chapter, I will summarise the findings of a literature review on OSCE implementation at two levels: international and national (SA), to comprehend the field of OSCE implementation and to develop appropriate research questions, aims, and objectives.

Chapter 2 Literature Review

2.1 Introduction

In this chapter, I conduct a scoping review to explore the literature available regarding the design and implementation of the OSCE. I explain why a scoping review approach is chosen, detail the scoping review framework, and illustrate the literature search strategy used. The literature is explored and discussed in two phases. Firstly, I describe how the OSCE should be implemented based on the key literature and global guidelines and then discuss some examples of OSCEs being implemented in new locations that have similarities that overlap with the Saudi context. Secondly, I discuss how OSCEs are being implemented in SA. All key literature concerning the use of OSCEs in Saudi medical schools is addressed, including all the extant publications that discuss the application of OSCEs in other health professions in SA.

2.2 Method: Scoping Review

In the following sections, I describe the scoping review approach, why I used it in this study, its process, the search strategy, and its results.

2.2.1 Defining the scoping review approach

There is no universal definition of a scoping review (Daudt et al., 2013). Multiple researchers have proposed definitions, but most commonly, it is described as an approach that systematically maps the literature of a particular field by exploring sources and collecting evidence to create conceptual clarity about a specific topic (Arksey and O'Malley, 2005; Davis et al., 2009; Levac et al., 2010). Although it is a relatively new literature review method, it facilitates mapping the literature and extracting data related to a particular topic (Peters et al., 2015; Maggio et al., 2020). A scoping review is a useful approach that enables researchers to understand the background and organise information to draw conclusions about what is and is not available in the literature on a specific topic (Anderson et al., 2008; Armstrong et al., 2011).

Using the principles of Arksey and O'Malley (2005) and Levac et al. (2010), this review is conducted for multiple purposes: (1) discuss, investigate, categorise and disseminate literature search findings related to OSCE implementation; (2) identify research gaps and explore the recommendations for future research in the field; (3) refine research questions and objectives; and (4) use the findings regarding the nature and methodology of the research in the field to inform current study.

2.2.2 Why the scoping review approach is used in this study

I chose the scoping review approach instead of other literature review approaches because it seems more closely aligned with the purpose of this project. In comparing the scoping review approach with the systematic review, I found that a scoping review would enable me to explore a range of literature relevant to the OSCE, identify gaps in the literature and refine my research questions (Levac et al., 2010; Armstrong et al., 2011). The systematic review approach was not chosen because it begins with a clearly defined research question and then critically appraises the literature (Moher et al., 2015; Suri, 2018). However, I had not clearly defined the research question before beginning the literature review, and my goal was not to evaluate the literature but to investigate the field.

A scoping review includes all relevant literature regardless of the research type (e.g., grey literature), whereas a systematic review is more focused on narrow parameters, like examining the quality of the literature as a basis for inclusion (Peters et al., 2015). Typically, systematic reviews evaluate the quality of included publications using indicators such as Shea et al.'s (2007) eleven quality indicators. Thus, standalone research can be produced using a systematic review (Arksey and O'Malley, 2005), but a scoping review is more useful for the current study because it allows for including all the relevant articles to clearly map the field and establish a firm background. For example, Saudi literature addressing OSCEs is very limited, which may be the result of a lack of research or a publishing bias in Western journals (Mulimani, 2019; Moriguchi, 2022). However, if the quality of the research is used as a basis for inclusion, many papers will be excluded, which may affect the richness of the literature review findings.

Moreover, a scoping review allows me to include studies that use different study designs, whereas a systematic review targets a pre-determined, specific study design (Gottlie et al., 2021). Systematic reviews tend to be more time-consuming than scoping reviews because they involve additional steps and are, preferably, conducted by a team of researchers, who, as a team, can devote more time to examining the quality of research designs. Hence, scoping reviews (1) tend to be more practical for an individual researcher who has a limited time-frame within which to complete the study and (2) are more suited to focusing on exploring and identifying gaps in the literature regarding a particular topic (Peters et al., 2015; Gough et al., 2017; Higgins, 2019).

2.2.3 Scoping review framework

This scoping review framework consists of six steps, which I have adopted and followed based on Arksey and O'Malley (2005), Levac et al. (2010), and Peters et al. (2015):

1- *Identify the primary research questions.* Based on the discussion in the Introduction Chapter, I am interested in addressing the following '*primary'* research questions:

a. How should the OSCE be implemented according to international guidelines?
b. How are OSCEs being implemented in countries that share some similar features with the Saudi context?
c. How are OSCEs being implemented in SA medical schools?
d. What are the opportunities and challenges facing OSCE implementation in SA?

2- *Identify relevant studies.* In this step, I identified the main elements that serve the research focus. For the sake of organising my literature review, I found it useful to explore the OSCE literature in two phases. First, investigate OSCEs' status and key publications in the international literature. Although OSCEs have primarily been discussed in Western contexts, exploring some non-Western countries that share similarities with the SA context is also useful. This phase sets the scene and provides the necessary background for this study. Second, in this phase, I have focused exclusively on OSCEs' status in SA. My primary interest is in how the OSCE is used in SA undergraduate medical schools; however, it would be beneficial to include all OSCE-related publications, even those related to other healthcare professions, because I believe that the topic is rarely discussed in the medical literature. Therefore, the study may benefit from the commonalities between different professions' implementation of the OSCEs in SA.
3- *Study selection.* In this step, I planned the search strategy and scope, which means specifying the search terms, techniques, and databases, as well as the inclusion and exclusion criteria, to best serve the research focus. This step is described in detail in the following section.
4- *Charting the data.* During the literature search, I highlighted the key points in each relevant paper. Then, I collated and summarised the extracted data in three tables (see Appendices 5, 6, and 7). In this step, it is important to maintain a log of all the literature relevant to my study to facilitate accessibility later. Furthermore, it enabled me to categorise the literature into groups and themes that discuss similar aspects of OSCEs.
5- *Reporting results*. In this step, I wrote the substantive sections. Summarising the key literature and relevant studies collected during the literature search allowed me to compare, contrast, interpret, and evaluate the identified publications. This step also enabled me to critically analyse information, support my argument and evaluate the significance and contribution of each piece of literature as it relates to my topic.
6- *Consultation.* According to Arksey and O'Malley (2005), this is an optional quality assurance step; however, I chose to include it because it would improve my review (O'Leary, 2017). Consultation in the current study involved requesting

comments and feedback from my supervisors to ensure that I had not missed any key literature or overlooked important aspects that are relevant to my research focus.

This process, as written, might appear linear. However, it is an iterative, time-consuming, and complex process of refining the research questions, adjusting the search strategy, reading/re-reading articles, summarising and tabulating information, and writing/re-writing the Literature Review Chapter.

2.2.4 Literature search strategy

The extant literature on the OSCEs is extensive. Therefore, I needed to focus the research process on the primary research questions so that I could gain depth without compromising the breadth that best serves the scope of this study. I identified all the possible terms related to OSCEs and their implementation around the focus of this study, which is undergraduate medical education. I specified the following search terms that are used to search academic databases (see Table 3):

- (Objective structured clinical examination) or (OSCE) in the title or abstract.
- (Implement*) was the truncation used to include (implement), (implemented) and (implementation) in the title or abstract.
- (Design*) was the truncation used to include (design), (designed) and (designing) in the title or abstract.
- (Quality) in the title or abstract.
- (Application) in the title or abstract.
- (Medic*) was the truncation used to include (medicine) and (medical).
- (Undergraduate).

Table 3. The formats used in the database searches.

#	Search terms
1	Search ((Objective Structured Clinical Examination[Title/Abstract]) **OR** OSCE[Title/Abstract])
2	#1 **AND** ((((implement*[Title/Abstract]) **OR** design*[Title/Abstract]) **OR** quality[Title/Abstract]) **OR** application[Title/Abstract])
3	#2 **AND** (medic*)
4	#3 **AND** (undergraduate)

I applied some inclusion and exclusion criteria during the stage of manual screening. This layer of filtration allows including only the publications most relevant to the research questions. The inclusion criteria utilised are as follows:

- Articles discussing the OSCE implementation, design, quality, or application process.
- Medical school settings.

- Undergraduate education.
- Published in English.

The second phase of this scoping review focused on OSCEs in SA, so the inclusion and exclusion criteria were modified to include papers discussing OSCEs in the following contexts:

- Undergraduate and postgraduate education.
- Any health profession.
- Saudi Arabia.

Hence, all published papers related to OSCE use in SA were included, which allowed conducting a comprehensive review of the literature related to this topic and benefited from identifying similarities in OSCE implementation in different health professions education.

I used the 'advanced search' option to search nine academic databases: PubMed, Ovid (for the following databases: Embase, Global Health, PsycINFO and MEDLINE), Web of Science, Cochrane Library, ERIC and Scopus. A manual search was also performed by manually analysing the key publications' reference lists to include any relevant articles the electronic search had missed. I included all the OSCE-related literature published since 1975, when Harden et al. (1975) first described the OSCE concept, until June 2023.

2.2.5 Conducting the search and charting the data

I began by searching for just the term 'OSCE' in all nine databases, which produced 11,667 results. Next, the pre-determined search terms listed in Section 2.2.4 were combined for additional searches using 'AND', which produced 674 results. The search strategy and the number of articles identified in each database are detailed and charted in Appendix 1.

In the manual references search, I identified 34 articles. Then, I performed three levels of filtration (as in Section 2.2.4) on the 674 database results and 32 manual exploration results (n=708 documents), which were all imported into EndNote referencing software. First, duplicates were excluded using EndNote. Next, the title and abstract of each article were screened and compared to the inclusion criteria. The table in Appendix 2 illustrates the reasons and number of publications excluded at this screening level. Then, the full text of these publications was screened to further assess whether they met the inclusion criteria. The table presented in Appendix 3 illustrates the reasons, number, and citations of publications excluded at this level. Finally, the full text of these publications was read and assessed for eligibility. The table presented in Appendix 4 illustrates the reasons, number, and citations of publications excluded at this level.

I utilised the preferred reporting items for systematic reviews and meta-analysis (PRISMA) (Moher et al., 2009; Moher et al., 2015) to depict the steps followed in the selection process (Figure 6. A PRISMA flowchart of the selection process.).

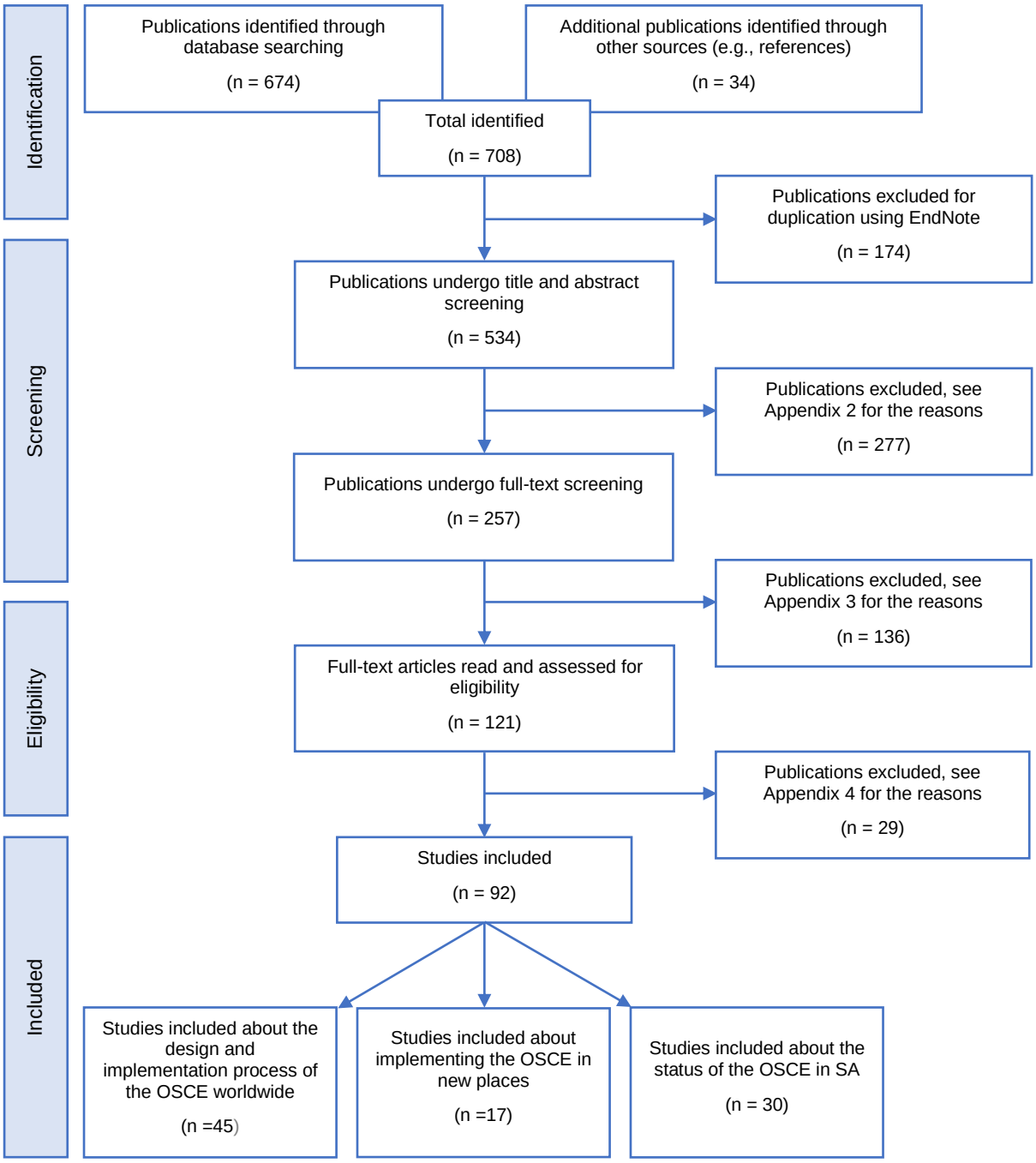


Figure 6. A PRISMA flowchart of the selection process.

I charted and summarised data from the included articles in Appendix 5 for the international publications, Appendix 6 for papers discussing OSCEs implemented in new locations, and Appendix 7 for publications specifically addressing SA. Charting data in

this manner facilitates accessibility to the relevant literature and provides an ‘at-a-glance’ overview of the publications included in each phase.

The literature search process identified an extensive number of articles, which demonstrates the effectiveness of the search strategy and search terms. Many of the articles identified are useful in discussing OSCE implementation. However, a considerable proportion of them is tangential to the current study’s focus. For example, some merely discuss students’ perceptions, and others focus solely on psychometric analyses. Furthermore, the electronic database search produced some irrelevant results, such as papers that mention ‘OSCE’ in the abstract as an example but address it minimally in the body of the paper.

Screening all papers and filtering out those that are only tangentially related to the topic of interest is not without benefit. Although it is a very demanding process, it helps me gain a sense of what is going on in the wider field, reveals the diversity in how researchers have addressed the topic and helps identify gaps in the literature. Furthermore, electronic searching is not without problems, as some irrelevant publications were included in the results, and some relevant publications that I found by other means were not included. For this reason, in the future, I need to remain vigilant for any relevant publications that were not captured electronically or manually.

2.3 Summary of relevant literature

The aim of this scoping review is to reveal not only what is available in the literature regarding OSCE implementation and what constitutes effective OSCE but also what is lacking in this area. Reviewing the literature, with this aim in mind, enabled me to identify potential gaps in the literature that this study could help fill (Peters et al., 2015). It also enabled me to identify the main sources and key elements of an OSCE. Structuring this discussion in two phases, the international and the SA-specific literature, allowed me to understand the status of OSCEs more deeply in each phase of this scoping review. I identified 92 publications for inclusion in this literature review. The first phase of this review discusses 45 publications focused on the OSCE implementation process and 17 papers focused on the use of OSCEs in new contexts. The second phase of this review discusses 30 papers addressing the OSCE in SA.

2.3.1 Phase one: Discussion of OSCEs in the international literature

The aim of this review is to identify how OSCEs should be implemented based on international literature and published guidelines. Since the focus of this study is on OSCE implementation in the SA context, in this section, I only demonstrate the key elements at a level of detail that enables me to make sense of and discuss how OSCEs should be implemented. Because addressing each element involved in OSCE implementation

could produce a lengthy literature review, keeping the aim of this scoping review in mind made it more practical, as well as useful, and paved the way for the second phase of this scoping review.

The OSCE implementation process comprises several elements, all of which should work harmoniously to produce a high-quality OSCE (Appendix 8). Therefore, although papers that discuss one element of OSCE implementation, such as those related to recruiting and training standardised patients (SPs), are useful, they do not inform this literature review as much as papers that discuss OSCEs as a whole and in an integrated way. The elements integrated to produce a high-quality OSCE are discussed in the literature in various ways, such as in books, guidelines, descriptive reports, and empirical studies.

2.3.1.1 Discussing the OSCE design and implementation process

In this section, I identify the key international literature that primarily focuses on designing and implementing high-quality OSCEs (Appendix 5). The following subsections seek to compile and logically arrange the main elements of OSCE design and implementation, as summarised in Appendix 8.

OSCEs are widely utilised in healthcare institutions around the world (Harden et al., 2015; Bearman et al., 2020). OSCEs are often considered to be the gold standard in the summative assessment of clinical competencies (Humphrey-Murto et al., 2013; Pugh and Smee, 2013). It is useful because it enables medical schools to ensure that their students can demonstrate the required clinical skills. However, OSCEs are applied in a variety of ways at different times (Heal et al., 2019) and sometimes, unfortunately, sub-optimally (Nyangeni et al., 2022). Hence, not all OSCEs are well-designed and implemented according to the aforementioned criteria (1.2.2) (Harden, 2015). Harden et al. (2015) assert that there are different levels of implementation of OSCEs, ranging from exemplary to inappropriate, and the latter can present a serious problem, such as passing incompetent students and failing those who are competent. They also argue that an OSCE is a powerful tool for assessing clinical competencies when administered properly. Therefore, institutions need to know whether their high-stake OSCEs meet the necessary standards for quality assessment (Pell et al., 2010).

According to Harden (2016), *'there are "good" OSCEs and "not so good" OSCEs. Reliability and validity are related to how the OSCE is implemented'* (p. 379). He calls a sub-optimally designed and implemented OSCE a *'potential OSCE (POSCE)'.* Heal et al. (2019) concluded, based on a survey of 19 medical schools in Australia, that there is wide variation in OSCE implementation between these schools. Therefore, ensuring the proper implementation of an OSCE is critical to ensure its quality (Carraccio and Englander, 2000; Gupta et al., 2010).

In the literature, the terms design, implementation, application, develop, structure, plan, and organise refer to the process of using an OSCE. However, I found the terms design and implementation are the most commonly used to describe this process and are overarching terms that could help someone comprehend the process. Therefore, I use these two terms throughout this study.

Although there is a general sense of what 'good' practise looks like in the OSCE literature, the criteria are rarely written into clear guidelines. It is incomplete in some areas or generic and too complicated for inexperienced assessors in other areas. Khan et al. (2013b) argue that the body of literature on how to implement a high-quality OSCE is limited. Nyangeni et al. (2022), in their literature review, concluded that there are limited records available in the literature to define and describe the quality of the OSCE. Researchers Selby et al. (1995), Boursicot and Roberts (2005), Khan et al. (2013a, 2013b), Pugh and Smee (2013), Zabar et al. (2013), Abdulghani et al. (2015), Harden et al. (2015), Daniels and Pugh (2018), Heal et al. (2019), Boursicot et al. (2020), and Bearman et al. (2020) provide a general picture of what an OSCE should look like, from a contemporary perspective. I found the structure of the AMEE guide for the OSCE by Khan et al. (2013a) to be a useful guide to organise this discussion as follows.

The OSCE planning committee, blueprinting and developing the stations

Designing OSCE assessments requires a great deal of effort and time, from blueprinting, station writing, organising, running the exam, and post-hoc quality assurance (Boursicot and Roberts, 2005; Bearman et al., 2020). Therefore, a team of OSCE experts and content experts should form an OSCE planning committee (Zayyan, 2011). The OSCE planning committee needs to assign OSCE organisers/teams a few months before the actual exam date to be responsible for the implementation process (Khan et al., 2013a). Their job is to monitor the implementation process' progress between the OSCE planning committee meetings, and report identified issues to the committee, which will make decisions on how to address them. However, some researchers report a lack of teamwork (Ataro et al., 2020) and faculty support to implement the OSCE (Khapre et al., 2020) hampered the OSCE implementation.

The OSCE planning committee also creates and agrees upon a blueprint matrix. Blueprinting involves sampling and deciding which skills an OSCE should cover from the curriculum and their individual weighting (Roberts et al., 2006; Abdulghani et al., 2015). The committee should assign a content expert for each station to be responsible for its written materials (Daniels and Pugh, 2018). Composing each station includes clarifying ideas and writing a station vignette, as well as writing candidate instructions, examiner instructions, and SP instructions if required, examiner and SP training (as discussed in the next section), and allocating time and equipment needed at the station (Casey et al., 2009; Pugh and Smee, 2013).

In each OSCE station, only one or a limited number of skills can be examined, enabling examiners to be very specific in what they assess (e.g., abdominal examinations), which is intended to ensure an OSCE is a highly structured examination. However, many researchers criticise OSCEs for compartmentalising how medical professionals approach patient care (Gupta et al., 2010; Nestel et al., 2011). Therefore, other clinical assessment formats (e.g., Mini-CEX and long and short clinical examinations) should be used to allow each format to complement the drawbacks of the other assessment formats (Newble, 2004; Casey et al., 2009; van der Vleuten et al., 2010).

There is no consensus on the optimal number of stations because it should be determined by the blueprint and the intersection of the number and duration of stations to determine the total time tested (Harden et al., 2015); however, Selby et al. (1995) suggest 10–15 stations, Gupta et al. (2010) recommend 12–30 stations, while Abdulghani et al. (2015) propose 12–16 stations. Nevertheless, it all depends on the purpose of the test, the amount of targeted curriculum content, the reliability required, and available resources. In general, the more stations included, the better it would be to, hopefully, ensure acceptable reliability (Brannick et al., 2011); however, due to feasibility issues, the OSCE planning committee should take a justifiable decision in this regard. In contrast, formative assessments should focus more on the educational and catalytic impact of the OSCE rather than the number of stations or duration of the examination (Latjatih et al., 2022; Al-Hashimi et al., 2023).

Similarly, the amount of time allocated for each station is not strictly agreed upon. Hijazi and Downing (2008) and (Harden et al., 2015) suggest 5-10 minutes, and Ware et al. (2014) consider 5–15 minutes to be adequate. However, Brannick et al. (2011) and van der Vleuten and Schuwirth (2005) suggest that longer overall testing times and larger numbers of stations are more important than the duration of each station. More importantly, the duration of each station should be decided by reaching a consensus between the material and exam experts, who compare the proposed duration with the nature of the stations, as well as ensure that all the stations can be adequately performed during the same amount of time (Daniels and Pugh, 2018). For example, task time alignment can be achieved by an expert faculty judgement on the time required for a competent candidate to complete the station (Pugh and Smee, 2013).

Fairness assurance and standardisation control

Standardising the OSCE testing environment is of paramount importance as it relates to implementation fairness and reliability. Therefore, the necessary arrangement of the examining venue should be taken seriously to ensure relevant standardisation. Standardisation could mean controlling all sources of error (Turner and Dankoski, 2008). Several variables significantly impact OSCE reliability and must be adequately controlled, such as examiner variation, patient (or SP) performance, question quality,

marking scheme, and station sample size (Turner and Dankoski, 2008; Khan et al., 2013b). High-quality OSCE assessments control and minimise the negative influence of these variables by ensuring standardisation, which can be achieved by providing proper training for OSCE designers, writers, examiners, patients, and SPs (Khan et al., 2013a; Fuller et al., 2017; Malau-Aduli et al., 2023). Briefing examiners, SPs, and students immediately prior to commencing the exam also helps keep certain and fair standards in action for all (Harden, 1990; Abdulghani et al., 2015; Gilani et al., 2021). However, excessive standardisation may have drawbacks, as it would result in inauthentic scenarios and SP delivery and reduce opportunities for more integrated, naturalistic stations (Yeates et al., 2022). Therefore, subject matter experts ought to strike a balance in OSCE standardisation to avoid excessive or insufficient standardisation.

To ensure an OSCE is as appropriately standardised as possible, the exam organiser needs to be available on exam day to supervise the stations and ensure integrity and quality control measures are implemented during the exam and in all parallel circuits (Daniels and Pugh, 2018). For the sake of exam security, when there are multiple circuits, proper arrangements may include quarantining later examinees in a suitable place and providing them refreshments due to the long wait times involved in this type of testing (Abdulghani et al., 2015). Nevertheless, the Ottawa conference consensus statement on the OSCE does not recommend quarantine per se (Boursicot et al., 2020), but it may be considered when the OSCE is poorly designed. However, in parallel circuits, such as multi-site OSCE, the OSCE designer must ensure appropriate design management for each site and conduct suitable post hoc analysis to guarantee that standardisation is maintained across sites (Pell et al., 2010; Harden et al., 2015).

According to Dyrvig et al. (2014) and Homer et al. (2020b), the scoring technique is a critical element of OSCE design and plays a central role in standardisation assurance as it guides the examiner's scoring process. Therefore, it should be set up carefully to reflect the important aspects of the skill being assessed and should provide for equitable scoring (Khan et al., 2013a). A marking scheme may be a checklist, a domain, or a behaviourally anchored rating scale, and either form may employ a global holistic rating (Hodges and McIlroy, 2003; Zabar et al., 2013; Setyonugroho et al., 2015). Some evidence suggests the global holistic rating has greater internal consistency and inter-station reliability when used in conjunction with a checklist than when used alone; however, its inter-rater reliability is lower than that of the checklist (Hodges and McIlroy, 2003; Patricio, 2012; Setyonugroho et al., 2015). Some researchers suggest that a rating scale is more valuable when the examiner is an experienced physician, but a detailed checklist is better for examiners who are not experts in the competency under examination (Turner and Dankoski, 2008; Gupta et al., 2010). However, with only a global rating, it would be difficult to use an advanced form of standard setting, such as the borderline regression

method (BRM), because global ratings are typically qualitative and lack numerical scores (Yudkowsky et al., 2019).

However, the global rating technique could work better for some types of stations. For example, it may work better in stations examining communication skills and behaviour (Pugh and Smee, 2013), but the checklist format might be preferable in stations testing specific techniques or precise, step-by-step clinical procedures (Abdulghani et al., 2015). Setyonugroho et al. (2015), in their systematic review of checklists used in OSCE, identified that one downside is the reliability and validity of checklists used to assess communication skills; they argue that there is no agreement on competencies required between checklists. Setyonugroho et al. (2015) and Heal et al. (2019) suggest developing and examining a standardised checklist for each competency domain. Although terms like checklists and rating scales are commonly used, their definitions are vague, and there is no consensus regarding their design and scoring (Homer et al., 2020b). For example, a checklist could include rating items. However, the choice and design of the marking scheme should depend on the examination's aims, the station's level of difficulty, and the intended level of assessment detail (Khan et al., 2013a; Wood and Pugh, 2019).

Examiner and SP training

Examiner characteristics and variation play a central role in OSCE reliability (Fuller et al., 2017; Yeates et al., 2019; Malau-Aduli et al., 2023), as unwanted examiner behaviours can increase bias and affect the assessment in many ways (Daniels and Pugh, 2018). However, according to Brits et al. (2020), some examiners are pushed to assess students in the OSCEs without proper training. For example, when the examiner asks questions beyond the pre-determined station instructions, it could affect the standardisation of the exam. Another example is examiner fatigue induced by watching multiple candidates repeat the same skill, which may be boring and repetitive and lead the examiner to become less attentive while assessing subsequent candidates, ultimately impacting the scoring accuracy and the exam’s reliability (Alnasser, 2016). Also, the effect of different examiners’ scoring behaviour being lenient or strict in parallel circuits of the OSCE or replacing examiners halfway through the same exam could affect the consistency of scoring in the affected station(s), as well as exam reliability (Fuller et al., 2017; Yeates et al., 2019). Therefore, considerable literature emphasises the importance of professional examiner training to minimise such undesirable effects (Boursicot and Roberts, 2005; Roberts et al., 2006; Gormley, 2011; Khan et al., 2013a).

Similarly, patient and SP training is essential to ensure accurate and consistent scoring (Gormley, 2011; MacLean et al., 2018). Recruiting real patients or SPs is challenging due to the need to ensure certain patients characteristic are present. The realism of SPs’ acting, the financial cost of hiring SPs, and their availability on the exam dates are

additional challenges (Collins and Harden, 1998; Zayyan, 2011). SPs could tire of repeating, for example, the same history, so they alter the signs and symptoms, which could attenuate the testing's fairness and reliability (Alnasser, 2016). Because SPs repeat the same scenario for each examinee, they may give clues or change behaviour when engaging with subsequent candidates (Khan et al., 2013a). Therefore, adequate training for SPs is important to ensure their roleplaying is standardised across all candidates (Groene et al., 2022).

Standard setting methods

Pass/fail decisions on high-stakes OSCEs must be carefully considered because the outcomes of summative assessment are either passing competent students, progression, and graduation of competent doctors to provide appropriate patient care, or failure and remediation (Ben-David, 2000; Pell et al., 2010). A primary concern is how to set station-level cut-scores and combine them to create an overall passing score, as well as how to handle students who receive borderline exam results (Norcini, 2003; Homer et al., 2017). Setting the cut-score could involve either criterion-referencing (also known as 'absolute') (i.e., based on pre-established criteria) or norm-referencing (also known as 'relative') (i.e., based on the performance of each cohort) (Casey et al., 2009). Several methods can be used for each type; for example, criterion-referenced can use Angoff, Borderline Regression, and Borderline Group methods, while norm-referenced can use Cohen and Wijnen methods (Ben-David, 2000; Norcini, 2003; Pell and Roberts, 2006; Yousuf et al., 2015). These standard-setting methods are widely described in the literature, but generally speaking, there is no right or wrong method; instead, the best method for the specific OSCE context should be selected (Gupta et al., 2010). However, the criterion-referenced method is more appropriate for the high-stakes OSCE, as test-takers' abilities and the exams' content difficulty change over time, rendering the norm-referenced method inappropriate for high-stakes assessment (McKinley and Norcini, 2014).

According to many authors, the borderline regression method (BRM) is a recommended approach to set the standard for the OSCE (Kramer et al., 2003; Dwivedi et al., 2020; Homer et al., 2020a; Park et al., 2021). However, It could, for example, be comfortable for some stakeholders to know the passing score before the exam (Norcini, 2003), but the BRM can be used to determine the passing score after all the students have finished taking the exam (Homer et al., 2017). Additionally, there are other variables to consider when setting the standards, such as whether each student must pass a certain number of stations or whether his or her average score on all stations should be considered (Homer, 2023). Nevertheless, the OSCE planning committee and the institute's assessment policy and regulations can, in combination, determine which method is more appropriate for the specific context based on logical justifications (Pell and Roberts, 2006). However, the literature indicates that many medical schools have an arbitrarily

fixed pass mark for the OSCE (sixty per cent in particular), which would not be defensible for high-stakes OSCEs (Tekian and Norcini, 2015; Rezigalla, 2016). Therefore, this practise is inconsistent with the literature's recommendation that each exam context should use the appropriate standard setting rather than a randomly selected standard setting (Homer and Darling, 2016).

Feasibility and logistics

As discussed in Section 1.2.7.2, the most cited disadvantage of OSCEs is the resource-intensive nature of this type of testing. Therefore, OSCE planning committees should consider this issue seriously and secure the required exam funding, venue and materials well in advance of administering an OSCE (Patricio et al., 2013; Badyal and Sharma, 2020). Some researchers have made suggestions for making OSCEs more feasible, such as establishing an OSCE station bank (Harden, 1990), collaborating with other medical schools (Zabar et al., 2013) and recruiting students as volunteer SPs (Casey et al., 2009). Harden et al. (2015) suggest sequential testing is an option to cut down on costs. This testing format involves examining all students on a small number of stations in the first stage and only administering the exam to borderline students during the second stage (Pell et al., 2013). However, this approach requires a reasonably robust psychometric analysis to identify borderline students, and it might not be feasible in some contexts (Homer et al., 2018).

Due to difficulties in allocating the necessary resources for the OSCE, some medical schools choose not to utilise it (Khapre et al., 2020). Another medical school with limited logistical resources documented their experience with the OSCE as an end-of-year exam, but staff and students identified numerous challenges, including a low number of stations, a lack of assessor training, and poor organisation (Ataro et al., 2020). To facilitate collaboration, several Australian medical schools formed The Australasian Collaboration for Clinical Assessment in Medicine (ACCLAiM) to share resources and experiences to facilitate the OSCE's implementation (ACCLAiM, 2023). This collaboration includes writing the stations, training the examiners, and exchanging expertise, all of which make OSCE implementation more feasible, with the ultimate goal of raising the quality of the OSCE across the country (Heal et al., 2019).

Quality assurance

OSCE quality can be investigated by conducting a qualitative and quantitative evaluation of the exam process (Tavakol and Dennick, 2011b). To perform a qualitative analysis of the exam, the exam organiser needs to obtain feedback from different stakeholders (i.e., examiners, examinees, participating staff members, and SPs) after the exam, which would provide valuable insight to rectify shortcomings and improve future OSCE administrations (Selby et al., 1995; Pugh and Smee, 2013; Zabar et al., 2013; Boursicot et al., 2022). Considering the students’ and staff members’ perspectives and perception

also has a positive impact on how they view the OSCE, which could enhance the students' educational benefit and acceptance of this assessment tool (Daniels and Pugh, 2018) and staff members' engagement on later exams, which could contribute to conducting successful upcoming OSCEs. Also, Khan et al. (2013a) suggest inviting expert external examiners as they can help check the quality of OSCE implementation and provide valuable feedback.

Quantitative analysis of the exam is also an important source of quality assurance (Fuller et al., 2012). Indeed, psychometric analysis can produce a range of data to ensure the validity and reliability of an OSCE (Tavakol and Dennick, 2011b). Pell et al. (2010) refer to evaluating the OSCE process and results as 'assessing the assessment,' and they emphasise the importance of using multiple metrics at the exam and station level to ensure assessment quality and reliability, as no single metric can accurately reflect exam quality. These metrics may include Cronbach's alpha, coefficient of determination (R^2), and between-group variation (Pell et al., 2010). Therefore, qualitative and quantitative analyses are both important to improve subsequent exams, teaching, and OSCE station banks.

Educational impact

The OSCE is not only an effective assessment tool but also an educational one (Abdulghani et al., 2015). The feedback provided to students after each station is highly recommended to improve the educational value of the OSCE (Harden et al., 2015). Many authors encourage examiners to provide feedback to examinees, which boosts the educational impact of an OSCE and, ultimately, the acceptability of this assessment tool (Casey et al., 2009; Pugh and Smee, 2013; Sterz et al., 2021). Feedback could be provided in many ways dependent on the stakes of the exam, e.g., immediately after each exam station, verbally or in writing, and handed to the examinees after the exam (Roberts et al., 2006; Khan et al., 2013a). According to Daniels et al. (2021) and Ngim et al. (2021), providing written feedback after the exam seems less stressful for the examinee and has greater educational benefits and more opportunities for further discussion between the examiner and the examinee. This feedback must be carefully structured to maximise its benefits, including what was good, what could be improved, and key action points (Boud and Molloy, 2013; Hattie and Clarke, 2018). Furthermore, using the OSCE as a formative exam can be an effective educational tool (Latjatih et al., 2022). Many authors reported the advantages of introducing peer-led formative or mock OSCEs, which can produce educational benefits while being very manageable in terms of logistics and staff workload (Moltner et al., 2020; Braier-Lorimer and Warren-Miell, 2022; Latjatih et al., 2022; Sharma et al., 2023).

To summarise, in this section, I have overviewed and discussed the main elements that need to be considered when designing and implementing the OSCE. I have extracted

pertinent information from the available key literature, which has allowed me to develop a thorough understanding of what constitutes 'good' OSCE design and implementation. The following section discusses how OSCEs are designed and implemented in different parts of the world, particularly in places where they have recently been implemented. Since some countries share similarities with the SA context, considering them informs the second phase of this scoping literature review, which focusses on the status of OSCE use in SA.

2.3.1.2 Implementing OSCEs in new contexts

This section explores the OSCE's implementation in countries that share a similar context as the SA, such as Egypt, Qatar, Iraq, Lebanon, and Pakistan or have recently adopted the OSCE, like Brazil, India, Taiwan, Haiti, Ethiopia, and Nigeria. Since most of the above discussion was derived from Western sources, investigating OSCE implementations in contexts that are similar to the SA context can inform this research on how other countries adopted the OSCE. Some medical schools have introduced OSCEs to their assessment systems more recently than others. I have identified eleven examples in different countries, described in seventeen papers describing the countries' experiences with implementing OSCEs. I expected that their experiences would inform the implementation of OSCEs in new places, such as SA. Appendix 6 includes a chart of the extracted data.

Abdelaziz et al. (2016) report their experiences with OSCE implementation in a limited resource context and conduct a follow-up study by Sallam et al. (2022) at Suez Canal University in Egypt. The University recently transformed from long-case traditional clinical assessments to a multidisciplinary OSCE that is conducted many times per year. They immediately faced the obstacle of balancing the OSCE's resource-intensive nature with the University's limited resources. The exam designer developed successful strategies to minimise the costs of administering the OSCE. For example, students are recruited to act as SPs and facilitators, and the exam is administered on the weekend to enable the full use of a whole four-floor building to run four parallel circuits of OSCE stations. They collected quantitative data regarding the implementation process by surveying students and assessors about their opinions and performed a psychometric analysis. Throughout the years, the main challenges were the ambiguity of some station instructions, the lack of SP training, and the inadequacy of station time (Sallam et al., 2022). However, the assessors and students reported overall satisfaction with the OSCE process in both studies. Students found it to be a fair and educationally valuable assessment tool, while assessors stated that it reflected the curriculum comprehensively and was useful for preparing students for clerkships. These findings regarding the students' perceptions are consistent with those of Shitu and Girma's (2008) study of

Ethiopian medical students, as well with the findings of a subsequent study conducted by Iqbal et al. (2009) in Pakistan.

Iqbal et al. (2009) report on a Pakistani medical school, which utilised the traditional subject-based teaching method, and the school's experience with reforming its OSCE in a resource-limited environment. They describe the previous OSCE as relatively passive (i.e., like an oral exam), which is consistent with Harden's (2016) description of this practice as poor implementation. Iqbal et al. (2009) state that the desire to reform the OSCE was motivated by two of the medical school's educators and had the dean's support. They were barely able to secure enough funding to develop a clinical skills laboratory to train students. In addition, the two medical educators conducted sessions to train other faculty members and SPs on their expected roles. Then, the new OSCE was developed, and the students' and faculty members' opinions were surveyed. Students regarded the new OSCE highly, praising its educational value and the post-OSCE feedback they received. Staff described the implementation of the new OSCE as a successful but resource-intensive process. This example demonstrates that despite limited resources, trained stakeholders with internal motivation and administrative support can enhance OSCE implementation. This highlights the significance of the leadership support of the dean in enhancing the assessment system, despite implementation challenges (Bryman, 2007).

Two papers from Qatar (Wilby and Diab, 2016) and Iraq (Alkhateeb et al., 2022) discuss aspects of the OSCE that are less relevant to the scope of this review since the first one was in pharmacy settings and the second one focused solely on student perspectives; however, they make some noteworthy points. First, both papers highlighted the lack of training for examiners and the fairness issue associated with examiner behaviour, so they propose additional calibrating and standardising of examiners. Secondly, SPs were difficult to recruit and train, so the authors of both papers suggest collaborating with other institutions to allocate suitable SPs and conduct training workshops to ensure the OSCE's fair conduct. Although these challenges are frequently raised in the literature, the Qatari and Iraqi contexts have a strong resonance with the Saudi context.

Troncon (2004), at the medical school of São Paulo University in Brazil, which adopts a traditional curriculum and introduces the OSCE for the preclinical third year, administered a questionnaire for three consecutive years to evaluate students' and assessors' opinions about the introduction of the OSCE. This study, the Pakistani study, and the Egyptian study all report their experiences in remarkable detail, reflecting their awareness of the requirements for implementing a high-quality OSCE. For instance, they invested considerable time and effort into training the examiners and SPs, explaining how they developed their station scoring and determined the pass/fail score. Furthermore, in the Egyptian example, the authors demonstrate the use of effective OSCE implementation strategies, from training staff on blueprinting and other tasks to

post-hoc quality analysis. However, Troncon's (2004) results reveal the students have an unfavourable perception of the new OSCE because they felt stressed by the exam and had a difficult time managing their time effectively while completing the stations. The assessors had a negative perception as they felt that the OSCE has limited educational value. For example, it compartmentalises real clinical scenarios, and the format is time- and effort-intensive to administer. After the first annual survey, São Paulo University's medical school amended its OSCE, but still the satisfaction rate remained unchanged over the next two years, so they decided to replace the OSCE with an assessment format that is easier to administer but, most likely, less standardised, without specifying which performance assessment tool they use now. Although the author acknowledged that the OSCE is a valid and reliable tool, he suggests it may not be compatible with 'traditional' medical schools.

Abdelaziz et al. (2016) calculated reliability using Cronbach's alpha, which, in their study, was 0.60, which is arguably below the acceptable level (Tavakol and Dennick, 2011a). Nonetheless, the authors did not discuss the quality control and comparative results between the parallel circuits to assess the reliability and fairness among them. However, a reliability assessment of the OSCE used at the newly established medical school at Lebanese American University reveals a very low Cronbach's alpha value of 0.43 (Karam et al., 2018). Likewise, Saeed et al.'s (2012) follow-up study of the aforementioned Pakistani medical school OSCE reveals low reliability of two of their OSCEs, with Cronbach's alphas of 0.53 and 0.48, which they link to the insufficient number of stations (10 stations) without providing further details. Furthermore, the Lebanese study shares few details to thoroughly investigate the lower reliability of the school's summative exam but indicates it utilises a small number of stations (7 stations) and is administered to a small number of students (53 students). According to the authors, these conditions could cause low reliability as it may be difficult to rely on statistical analysis results with such small numbers (Karam et al., 2018). They also identify a teaching problem based on the high failure rates on two stations. Therefore, whilst the first OSCE implementation might not always be successful, these examples highlight the importance of conducting post-hoc evaluations, which highlight the weak points in the OSCE implementation process, as well as issues related to the teaching process and opportunities for improvement.

The drawback in these four examples is that they all depend exclusively on quantitative measurements, which, in isolation, cannot precisely evaluate the implementation process. This could be overcome by combining them with qualitative analysis to capture the causes and investigate the issue more deeply (Tavakol and Dennick, 2011b). Thus, accurate implications cannot be drawn from merely numerical data.

However, Abdelaziz et al. (2016) suggest a possible reason the Egyptian OSCE implementation achieved high satisfaction among students and staff is that they adopted the problem-based learning approach to teaching. The Brazilian medical school uses

traditional teaching methods, and Troncon (2004) suggests OSCE implementation is not ideal with traditional teaching methods as OSCEs require a supportive educational environment. This point has been raised in SA as Bajammal et al. (2008) postulated a relationship exists between the curriculum design, teaching method, and successful use of an assessment format. However, the Pakistani medical school also used the traditional teaching approach and had a better outcome than the Brazilian example. The Egyptian study mentions another reason for their successful implementation of the OSCE, which is also more in line with the Pakistani case: The administrative faculty supported the OSCE implementation, and the academic staff were enthusiastic and ready for change and improvement despite the resource shortage. Troncon (2004) indicates the crucial role of educational administration in inducing supportive culture changes that promote the desire among faculty members for constructive change, which is related to the introduction of change discussed in Chapter 7. This factor seems to be a plausible causal factor in their experiences with the OSCE. Similarly, some Saudi publications emphasise the strong positive impact of administrative support and motivated staff members in implementing a successful OSCE (Ware et al., 2014).

Bhatnagar et al. (2011) report their experiences with a newly implemented OSCE in an Indian medical school and conclude that because the process is time- and effort-intensive, staff members need to maintain their motivation in the OSCE development and implementation process. They raise a similar idea that Troncon (2004) raised regarding the need for staff members to be internally motivated and believe in the usefulness of the OSCE, as well as external motivation, which could be created and maintained by a supportive institutional culture. Bhatnagar et al. (2011) suggest, after their implementation experience, that the OSCE can be valid, reliable, and feasible in the Indian context and, hence, encourage its dissemination in other Indian medical schools. However, in a more recent Indian study, Khapre et al. (2020) investigated the faculty's perspectives on OSCE implementation at their medical school. Based on their analysis, they conclude that there is some resistance among the administrative and academic faculty to implementing this new assessment tool; unsupported policies, limited resources, and the absence of a faculty training programme are also obstacles to the OSCE's implementation.

Kaohsiung Medical University, in Taiwan, demonstrates the role of internal and external motivation in the development of their OSCE (Tsai et al., 2008). The motivation to implement this tool was initiated and supported at their institution since its first introduction in 1992. The introduction of the OSCE earlier than other medical schools, combined with faculty motivation, established them as the leading OSCE centre in Taiwan, from which other medical schools can learn (Huang et al., 2007; Lee et al., 2008; Tsai et al., 2008).

In some countries, the OSCE is considered a totally new assessment tool, such as in Haiti and Nigeria. In Haiti, Sainterant et al. (2019) briefly describe how the OSCE was first introduced there with the assistance of two Canadian doctors who volunteered to train the staff of a Haitian medical school and assist in the implementation process. The authors provide very few details in their report. However, they mention there was a lack of funding for this assessment, which they overcame by asking students to play the role of SPs voluntarily, but they describe it as a positive experience overall that benefited all stakeholders. In a Nigerian medical school, Onwudiegwu (2018) introduced the OSCE recently and developed guidelines for OSCE that, he suggests, could help other medical educators in African who want to implement and benefit from the OSCE. Instead of demonstrating or evaluating his recent experience with the OSCE, Onwudiegwu (2018) opted to encourage OSCE use, as it is not a common assessment tool in the African context, by producing a concise and brief guide to facilitate OSCE implementation in other African medical schools. However, another African author discusses their OSCE experience at an Ethiopian university (Ataro et al., 2020). Although examiners and students found this new assessment tool to be fair and useful, they critiqued its poor organisation and the insufficient number of stations. Therefore, the authors suggest more teamwork in OSCE planning and preparation to overcome the challenges. These examples demonstrate educators' awareness and appreciation of the OSCE and its flexibility in different contexts.

Therefore, context is central when considering OSCE implementation. The context may include local institutional culture, assessment practises, regulatory requirements, faculty expertise, educational settings, and resources. Besides carefully considering the OSCE elements explored at the beginning of this phase of the scoping review, utilising lessons learned from examples of recent OSCE implementation in similar contexts can be helpful in implementing a high-quality OSCE. Some of the examples discussed can be useful for discussing the OSCE implementation in SA, as there are some similarities.

2.3.1.3 Summary of phase one

I found Patricio et al. (2009) conclusion, in their wide-ranging OSCE review, to be accurate, as they identify 104 papers reporting OSCE studies, the vast majority of which suffer from a lack of key logistical information, making interpretation very difficult and, ultimately, lowering their research value. For example, The Best Evidence Medical Education (BEME) was unable to include many OSCE publications in their systematic review due to a lack of detailed information (Patricio, 2012). The majority of OSCE literature consists of descriptive papers about OSCE implementation outcomes but not the OSCE design and implementation process (Duerson et al., 2000; Dadgar et al., 2008; Al-Haddad, 2009). A considerable proportion of the literature is focused on post-test analysis and does not provide enough explanations or details of the implementation

process to understand how they impacted the analysis (Rekany and Aldabbagh, 2010; Mondal et al., 2012; Tijani et al., 2017). Some papers discuss student perceptions rather than describing OSCE design and implementation (Pierre et al., 2004; Rehman et al., 2012; Mamatha et al., 2017). Moreover, these papers only illustrate their study methodology and results without providing adequate background information regarding how they designed and implemented their OSCE, which prevents the reader from understanding the implementation process well enough to draw conclusions regarding relationships between the actual implementation and the studies' results. Although these papers informed the research by providing good background information for various OSCE practises in different countries, their value is limited, given my research focus on understanding the implementation process. These types of papers are less useful in helping assessors develop high-quality assessments (i.e., they provide no clues as to whether what they have done during the design stage constitutes good or poor practise). For additional information regarding the excluded papers, see Appendix 3.

This scoping review demonstrates that the OSCE can be valid, reliable, fair, and educationally valuable. The literature suggests that the OSCE is beneficial, and most stakeholders are pleased with it as a performance assessment instrument. However, several papers identify challenges to OSCE implementation, such as low support from some administrative and academic faculty, unsupportive policies, a lack of examiner and SP training, limited resources, a small number of stations, and a short duration per station. Nonetheless, these enabling and impedimental factors are highly dependent on the broader context where the OSCE is implemented, as well as the institutional culture.

In general, there is limited information about what good design and proper implementation look like. Additionally, limited research has been conducted on adopting best practises for designing an OSCE and using multiple metrics to ensure the quality of its results and analyses (Pell et al., 2010; Daniels and Pugh, 2018). Although a well-designed OSCE can meet most of the aforementioned criteria for an effective assessment (Sections 1.2.2 and 2.3.1.1), there are few well-defined guidelines for designing and administering a high-quality OSCE exam in different contexts. However, there are some overarching criteria and general guidelines for all assessments that can assist in designing a high-quality OSCE (see Figure 7. The factors that contribute to the OSCE quality., modified from the work of Pell et al. (2010) and Harden et al. (2015)). However, available criteria and guidelines may not be useful for inexperienced OSCE designers and examiners, so it may require assessment specialists to utilise them. The next section will explore the OSCE's status in SA and discuss publications that are specific to the Saudi context.

Figure 7. The factors that contribute to the OSCE quality.

2.3.2 Phase two: Discussion of the OSCE in the Saudi literature

This phase of the scoping review aims to investigate how the OSCE is being implemented in SA specifically. The first phase comprised an overview of the status of the OSCE globally, which helped in comparing and contrasting it with the current status of the OSCE in the Saudi literature. In the previous sections, I showed the main and common characteristics that could be transferable to different settings during OSCE implementation. However, context is crucial since the OSCE design would be largely influenced by the environment where it is implemented (Hodges, 2003b). Thus, the characteristics of OSCE implemented in other countries may not have the same results if the same process is used for the OSCE in SA (Wilby et al., 2016). Likewise, obtaining the same OSCE within one country is difficult due to the diverse contexts of medical schools. The 41 medical schools in SA exhibit diversity in their resources, learning approach and capacity (Bajammal et al., 2008).

I have identified 44 publications that discuss the OSCE in SA. This includes all publications in Saudi literature that mention the OSCE and Saudi in the title or abstract and pertain to any health profession or level of education. I have charted the main data extracted from all these papers and classified the articles into three categories according to their relevance to my research focus—useful, little benefit and not useful (‘excluded’) (Appendix 7). The numbers in each category were 4, 26 and 14 papers, respectively. The 30 included papers are from undergraduate medicine settings (23), postgraduate medicine (6), and pharmacy (1).

Literature about the OSCE in SA is limited, given that the oldest medical school in the country was established in 1967, and more than half of the country’s 41 medical schools have existed for 12 years or less (MOE, 2022). As a result, the literature on medical education in the Saudi context is limited. Saudi literature on assessment, and OSCE in particular, is even more limited.

The first application of the OSCE in SA was in 1987, as documented by Abdullatif (1992) (Appendix 9). The next publication about OSCE is by Ismail (2001), who aims to capture students’ opinions about it; next, Khairy (2004) briefly reports his opinion on the OSCE's feasibility and acceptability in SA. The other 41 papers were published later. This reflects the fact that the OSCE is a relatively new assessment tool in Saudi medical education (Alsaid and Al-Sheikh, 2017).

2.3.2.1 Status of the OSCE in SA

I have included 30 papers for discussion in this phase and excluded 14 of them for the reasons stated in Appendix 7. I have structured this section to discuss the Saudi OSCE literature concerning this study's focus on the implementation of the OSCE. I found the framework of good assessment by Norcini et al. (2018) to be a useful framework to

organise this discussion. Therefore, I have organised it as follows: discussing implementation, validity, reliability, feasibility, stakeholders training, educational impact, and acceptability.

OSCE implementation

The papers most relevant to my primary research questions involve the postgraduate setting (Hijazi and Downing, 2008; Ware et al., 2014), not because of their context but because they discuss elements of the OSCE implementation process. Hijazi and Downing (2008) aim to encourage postgraduate programmes to adopt the OSCE and facilitate its use. They offer a useful descriptive guide for implementing the OSCE, but it appears too brief and lacks certain aspects pertinent to the undergraduate setting, such as providing students with feedback and the logistical resources of medical schools. The other descriptive manual available in the Saudi OSCE literature is recommended by the Saudi Commission for Health Specialities (SCFHS), developed by Ware et al. (2014). This short manual is directed to the postgraduate training centres in SA to help them plan and deliver the OSCE for postgraduate health speciality trainers. Both publications provide a good overview and overarching principles for designing the OSCE, which are in line with the key resources discussed above (2.3.1.1). However, they are not specific to medicine or undergraduate education; rather, they are more applicable to medical residency programmes. In general, they discuss blueprinting, item construction, exam logistics and recommendations for running the OSCE on the exam day. Their main and relevant interests are in implementing highly standardised, well-structured, and fair exams in hospital settings. The other relevant point these guides emphasise is the need for training for all stakeholders, such as examiners, writers, examinees, coordinators, and SPs. However, Hijazi and Downing (2008) and Ware et al. (2014) do not address acceptability issues, post hoc analysis, educational value, or how to provide feedback to students, indicating that these documents are not comprehensive manuals as they claim to be. However, they may rely on triangulating evidence from different sources to support implementation, as creating a comprehensive directive guide may be challenging. However, their prominent direction is to examine residents for accreditation purposes, which is not the focus of this study.

Bajammal et al. (2008) and Abu-Zaid et al. (2020) raise the need for assessing new medical graduates in a national assessment that includes an OSCE part, while Hijazi and Downing (2008) recommend assessing the current residents with the OSCE instead of merely long case assessment or end-of-rotation rating. This reflects medical educators' awareness of the value of the OSCE as a valid and standardised assessment in medical education in SA. Both, Bajammal et al. (2008) and Abu-Zaid et al. (2020) do not discuss the implementation of the OSCE; rather, they argue for its need to be used in the national exam.

Validity

Five papers have tried to assess the validity of OSCEs (Abdullatif, 1992; Bakhsh et al., 2009; Amr and Amin, 2012; Moeen-uz-zafar et al., 2015; Inayah et al., 2017). Abdullatif (1992) calculates the concurrent validity (i.e., criterion-related validity) of the OSCE with the final traditional long-case clinical examination of four groups of fifth-year medical students to find a strong positive correlation (Spearman coefficient r = 0.76). This result is for the first known OSCE in the Saudi literature, but other than the result, no details about the OSCE design are mentioned.

Due to the concerns among the medical school faculty at King Abdulaziz University about the validity of the OSCE, which was introduced for the first time in 2003, Bakhsh et al. (2009) conducted a validity study to assess the validity of their OSCE. However, they examine the validity of their OSCE without using any validity framework, instead comparing the results to those of other school exams. One of the major concerns is the fragmentation style in examining patients using the OSCE method. The authors find a strong positive correlation between students' OSCE scores and their global scores on all assessments (Pearson correlation r = 0.78). This study supports the faculty and encourages them to use the OSCE more widely. Similarly, Amr and Amin (2012), from the College of Medicine at King Faisal University, claim a significant correlation between their Psychiatry OSCE and the traditional oral examination, but with no clear explanations for this claim. At the College of Medicine of Qassim University, Moeen-uz-zafar et al. (2015) analysed the correlation between the OSCE and MCQ final examination results of fourth-year students in the internal medicine course, finding a moderate positive correlation between the two (r = 0.62). These three studies show the concurrent validity of their OSCEs, which might indicate their good implementation. All studies have not explained how the OSCEs were conducted or considered any implementation-related issue, making their research less beneficial for my study than it would otherwise have been, but the studies still provide a partial picture of their concurrent validity.

According to Inayah et al. (2017) at Alfaisal University, they administered a well-developed short mid-semester OSCE 12 weeks from the beginning of the semester, which resulted in a good correlation with the final long OSCE held a month after the first one. In this study, the authors also estimate a high predictive validity of the mid-semester with the final OSCE. In their conclusions, the authors suggest that a well-implemented short OSCE (three stations) can be a good alternative to a long OSCE (five stations) in low-resource settings. However, this is not a highly beneficial study to me since the authors do not describe the process of their implementation to enable us to judge whether it was well developed or to reproduce the study elsewhere. Furthermore, the number of stations in each OSCE exam seems inadequate for producing a valid or reliable exam (Khan et al., 2013a). In summary, from this scoping review in the Saudi

context, no single publication shows a complete validity evaluation showing the blueprint or station sampling as a source of evidence. Moreover, these studies rely on classical or '*granular'* definitions of validity rather than the more inclusive contemporary approaches set out by Messick (1989) or Kane's (1992) validity argument, as described by Boursicot et al. (2022).

Reliability

Two papers by Al-Naami (2008) and Al-Naami et al. (2011) discuss their OSCE results analysis. At a well-established medical school of King Saud University, the oldest and best-funded medical school in SA, Al-Naami (2008) assesses the validity, reliability and feasibility of their OSCE, which was first introduced in 2005. Three years later, Al-Naami et al. (2011) report new results on their OSCE and the interventions they employ for the same end-of-year surgical exam. The staff and students judged the 2008 OSCEs' face and construct validity results to be very good, and this remained the same for the 2011 OSCE. In 2008, Pearson's correlation coefficient was used to calculate the OSCE's correlational validity with the MCQ exam; the result was 0.5; in 2011, this value improved to 0.65. The internal consistency of the 2008 OSCE consisted of 24 rated stations and lasted for 160 minutes, and its Cronbach's alpha was 0.73, reflecting moderate reliability. In contrast, the 2011 OSCE had only 15 rated stations for the same exam, and it lasted 120 minutes. This time, the exam had a higher reliability: Cronbach's alpha was calculated to be 0.8. In 2008, the author reported that the staff rated the exam feasibility as poor because it consumed a notable amount of time, significant effort, and many SPs. However, in the later assessment, the authors reported better feasibility.

The improvement in all sources of evidence to their OSCE quality indicated the successful interventions that they employed. The authors attribute this to their medical school senior administration efforts in focusing more on examiner training, changing a checklist to be a global rating, and increasing staff awareness of the significance of this assessment style. These measures all enhance the internal motivation to contribute actively to this exam; as Ware et al. (2014) describe, this is one pillar for successful OSCE implementation. The moderate internal consistency of the 2008 exam reflected an adequate number of stations, but this also made the exam less feasible as it required many examiners, SPs, and other resources. In contrast, the fewer stations in the later exam may have made the exam more feasible. This may imply that the effect of examiner and writer training workshops and their internal motivation, as reported in the 2011 OSCE, is more effective in enhancing exam reliability than the mere length of the exam. Comparing the two examples suggests that better implementation can make the shorter exam (the 2011 exam) more reliable and feasible than the longer exam (the 2008 exam). These examples emphasise the importance of good implementation and the exam being acceptable to stakeholders who are well-trained to implement a successful OSCE.

In both papers, the authors describe their OSCE method reasonably well, but they provide little supporting evidence for the choices made during the preparation and introduction of the exam. However, they indicate the importance of the training sessions that they provide for the staff.

Feasibility

Similar to the international literature, SA has a feasibility issue posed by the complexity of the OSCE's implementation. It is a demanding assessment that needs the allocation of considerable labour, time, effort, and space (Khairy, 2004; Bakhsh et al., 2005; Al-Naami, 2008; Elfaki et al., 2008; Badawi et al., 2023). One paper by Elfaki et al. (2008) focuses on seven years of implementing OSCE in one medical school and details some challenges they faced in their OSCE implementation, including resources being exhausted for examiners preparing and running the exam, the demand for patients or SPs, and the lack of standardisation in some OSCE stations. While they claim their OSCE proved successful, they say little about how they implemented it in the first place to support their claim. Since the OSCE requires a significant amount of resources and effort, Elfaki et al. (2008) emphasise the need for the commitment of all stakeholders to maintain and develop the OSCE implementation.

A report published by Bakhsh et al. (2005) provides significant details on the logistic aspect of the OSCE on the day of the exam, concluding that it is a resource-intensive assessment format that requires extensive funding and staff commitment to run it successfully. Although Al-Naami (2008) reports from a well-established college of medicine, he raises this issue; nonetheless, they seem to afford its expenses better than other newly established or less funded schools do. At one teaching hospital, Ismail (2001) conducted a summative OSCE for the neurology course of fifth-year students. There were only two stations, which seems invalid and unreliable and provides no opportunity for proper blueprinting (Abdulghani et al., 2015); however, the author discusses the students' positive perception rather than this crucial validity issue. Ismail (2001) does not explain the reasons for this low number of stations, but it could be due to limited resources or limited awareness about the basics of the OSCE. Such exams seem to be resource-wasting since they have insufficient validity and reliability, which hampers their usefulness.

Examiner and patient training

Al-Naami (2008), Hijazi and Downing (2008), and Ware et al. (2014) discuss the value they could gain from training the examiners and SPs, which would affect the validity, reliability and acceptance of the exam, as well as the preparedness for this type of assessment. Training helps maintain standardisation and fairness across the exam stations and circuits, which is one of the basic notions of the OSCE. However, it is hard to find any other publications mentioning its importance in Saudi literature, so it appears

that this concept is not gaining its deserved attention. For example, Elfaki et al. (2008) published their 7-year experience with the OSCE, claiming it can work as a concise guide for newcomers to the OSCE. In this work, the researchers totally neglect essential issues, such as the importance of training the examiners, SPs, the criticality of blueprinting, and the need for conducting psychometric analysis. This may show the amount of deficiency in this field in the local literature and practice.

Real patients can become tired and uncooperative, which attenuates the exam's fairness and standardisation (Elfaki et al., 2008; Bakhsh et al., 2009). This is one of the problems that has been underestimated about the OSCE implementation in SA as there is limited awareness of the need for trained SPs as a beneficial factor for standardisation and feasibility, not to mention SP training and medical schools' collaboration in this regard. However, this significant issue has only been raised in a few publications (Al-Naami et al., 2011; Alsaid and Al-Sheikh, 2017).

Educational impact

According to OSCE global literature, the consensus is that the OSCE is educationally beneficial (Khan et al., 2013b; Harden et al., 2015). However, Alghamdi et al. (2016) find that students do not consider the OSCE to be valuable for their educational process; the authors do not comment on this viewpoint, but the overall students' perception of the OSCE in that study is negative. In contrast, the students' opinions in the questionnaire studies by Ismail (2001), Alaidarous et al. (2016), Alsaid and Al-Sheikh (2017), and Badawi et al. (2023) identify the OSCE as educationally valuable; in the studies the assessors discovered the areas that need to be revisited and rectified in future teaching. Alsaid and Al-Sheikh's (2017) study involves a survey of student and staff perceptions about the OSCE at a well-established medical school. Interestingly, the school only adopted the OSCE in 2013 to replace long case traditional clinical assessment. In the results, both students and faculty provide positive comments, identifying the OSCE as an educationally beneficial and fair assessment tool. Yet, the authors' discussion is somewhat superficial, and there are not enough details provided to enable readers to interpret the findings meaningfully.

However, the educational impact of OSCE could be enhanced by implementing high-quality OSCEs and following the recommendations of the pioneers in the field, as discussed above in this scoping review (2.3.1.1); for example, this could involve providing students with written qualitative feedback in each station (Harden et al., 2015). For example, Alaidarous et al. (2016) administrated 12 formative OSCE stations for residents in a university hospital, and the examiners provided feedback for each candidate. The residents highly appreciated the 3 minutes of feedback after each station and even suggest increasing the time for later exams. The candidate responses to this experiment were strongly positive; thus, the authors and residents advocate for the

continued use of this examination and the incorporation of this modality into the undergraduate curriculum.

Furthermore, in their study, Badawi et al. (2023) conclude that the OSCE can be utilised as a teaching tool. They show the students videotapes of clinical skills performed in some OSCE stations and ask them to assess them using checklists similar to those used by the examiners in the OSCE stations. After this intervention, the authors measured the students' understanding and information retention by conducting some applied knowledge tests, concluding that the OSCE is a useful teaching method. These examples suggest the promising educational value of the OSCE in the Saudi context. These findings support the notion that when there is a focus on integrating feedback, as with any good assessment, this is likely to drive the value and impact of the OSCE (Hattie and Clarke, 2018; Ngim et al., 2021).

Acceptability and students' opinions

Many of the papers (12 papers out of 30) in the SA OSCE literature discuss the issue of students' perception and acceptance of this assessment format. Saudi students' perceptions are in favour of OSCEs, which is largely in line with the global literature, such as the studies of Griesser et al. (2012) and Pierre et al. (2004). From questionnaire responses, it seems that most students in SA are in favour of the OSCE as an assessment tool (Ismail, 2001; Khairy, 2004; Bakhsh et al., 2005; Raheel and Naeem, 2013; Ibrahim et al., 2015; Moeen-uz-zafar et al., 2015; Al-Eidan et al., 2016; Alaidarous et al., 2016; Elfaki and Al-Humayed, 2016; Alsaid and Al-Sheikh, 2017; Alsulimani et al., 2020; Ansari et al., 2021). However, using questionnaires to capture students' perceptions about this assessment format might not allow full investigation of the reasons behind these views compared to qualitative methods.

Nevertheless, the results of prior studies indicate the high acceptability and satisfaction with OSCEs among students. Khairy (2004), Al-Eidan et al. (2016) and Alaidarous et al. (2016) find that the students were mostly pleased with the organisation and smooth running of the OSCE; Moeen-uz-zafar et al. (2015) indicate that the students found OSCEs useful for their learning progress; and Alsaid and Al-Sheikh (2017) report that the students viewed their OSCE as being fair and better compared with the traditional clinical examination. These views reflect that students' judgement largely depends on how the OSCE is constructed and introduced. However, this could be a premature judgement since none of the above studies gives useful details for their design and implementation of the OSCE, which would enable the readers to understand the source of satisfaction. Nevertheless, it is possible that the assessments are of a lower standard than might be expected, allowing students to perform well on such exams, which may influence their opinions. From another perspective, nine studies introduced their questionnaire immediately after the OSCE, but the other three do not mention the

questionnaire administration time. Usually, after long exams, students are exhausted, and may wish to avoid upsetting their faculty members before the grades are released. Both considerations may attenuate the validity of their responses.

Although Alaidarous et al. (2016) describe some aspects of the process of their OSCE implementation, they do not discuss the validity assurance procedure or feasibility of their exam. Alaidarous et al. (2016) conducted their OSCE at a well-funded hospital, which may explain why they did not discuss feasibility in detail. Their study focusses on a formative OSCE for internal medicine residents. Participants completed a questionnaire followed by three open-ended questions to obtain some qualitative feedback from residents. The residents' opinions reflected positive attitudes and acceptance of this formative assessment. Alaidarous et al. (2016) report the importance of the timely feedback given to the residents in the OSCE, in contrast to most papers published about the OSCE in the SA, which do not discuss the significance of the feedback given to candidates.

On the other hand, Alghamdi's et al. (2016) survey from one of the well-established medical schools of King Abdulaziz University explores student perceptions about their experiences with the OSCE. The results reveal that students have concerns about the OSCE's reliability and fairness due to inter-evaluator and inter-patient variability, as well as stating that too little time is allocated for each station. More than half of the students perceived that the OSCE is stressful and does not reflect real-life situations, so they questioned its educational impact. They indicated issues in the design of this exam, for example, a mismatch between the skill to be performed and the time allocated for it, which may lead to incomplete performance and induce stress. The authors conclude that these results contradict those of another study conducted at the same institution by Bakhsh et al. (2009), who advocate and encourage the use of the OSCE and demonstrate its reliability in their study findings. Therefore, Alghamdi et al. (2016) conclude that they need to scrutinise how they conduct their OSCE.

Several studies demonstrate that students view the OSCE as a stressful examination that causes test anxiety, which could lead to underperformance (Raheel and Naeem, 2013; Moeen-uz-zafar et al., 2015; Alaidarous et al., 2016; Alghamdi et al., 2016; Hadi et al., 2018; Arain, 2021). Although the researchers are aware of the positive effect of a certain level of stress on exam preparation, they report a negative overall influence of stress on students' performance. In Hadi et al.'s (2018) study, pharmacy students rated the OSCE as more stressful than their written assessment. This may have been due to their awareness of the 'critical element', which the authors define as the element in the exam station that can result in a student's failure if their performance does not address it correctly. The authors claim that students perform comparable tasks in formative OSCEs; however, this does not appear to be effective in alleviating students' anxiety.

Therefore, they recommend further investigation of the source of stress before attempting to administer OSCE stations.

Other stress factors are indicated by Moeen-uz-zafar et al. (2015) and Ansari et al. (2021), including the lack of time allowed to execute the skills at most of the stations. For example, students take a long time to read the station scenario and directions; they should be shorter and clearer to allow more time for focussing on executing the task. According to the students' comments, when OSCEs are administered in English to Arabic-speaking students, the language barrier can induce anxiety and affect the accuracy of their responses (Alghamdi et al., 2016; Alnahdi et al., 2021; Ansari et al., 2021), as these issues have the potential to undermine the OSCE's validity, they should be seriously considered by OSCE designers to mitigate them further.

From the discussion of the status of the OSCE in SA, I have identified some opportunities that would support the development and implementation of the OSCE to a high standard, as well as some challenges that medical schools in SA must overcome to implement the OSCE to a high standard. The subsequent two sections discuss these facets.

2.3.2.2 Reflection on opportunities and challenges for improvement of the OSCE

Opportunities

From this scoping review, my observation is consistent with the findings of Bajammal et al. (2008) in that there is a lack of consistency in assessment; OSCEs are designed and implemented in different ways across Saudi medical schools. These differences may be due to the lack of guidelines and frameworks on how to implement high-quality OSCE. In contrast with the two concise and non-inclusive OSCE guides that direct postgraduate OSCEs for accreditation purposes (Hijazi and Downing, 2008; Ware et al., 2014), I cannot locate a single OSCE guide for undergraduate education in SA. Although the process of developing the OSCE in undergraduate settings may be similar to that of postgraduate, the purposes, needs, resources, and settings of undergraduate assessment are different. Therefore, there is a good opportunity for medical schools to collaborate in developing a set of recommendations that can guide OSCE implementation and enable them to conduct high-quality OSCEs. This could enable their graduates to achieve superior performance on the national OSCE exam that was proposed to the SCFHS to ensure the competency of medical practitioners (Abu-Zaid et al., 2020; SCFHS, 2023). This is an excellent opportunity for medical schools to highlight their assessment excellence, assure stakeholders of the quality of their graduates, and prepare them to perform well on post-graduate exams, giving them a competitive advantage in the labour market.

Weak evidence regarding the presence of OSCE guidelines and structured frameworks for undergraduate settings in the local context provides medical schools with the opportunity to innovate and adapt the OSCE to different contexts. This allows assessors to develop it in their own ways; however, at present, they lack the necessary educational evidence to do so appropriately. Therefore, context-relevant evidence must be established to ensure the OSCE's validity.

The literature suggests ways to improve the OSCE in SA, such as introducing training courses for all OSCE implementation staff and establishing collaboration between medical schools to share OSCE station banks, examiners, SPs, expertise, resources, and even well-equipped venues (Elfaki et al., 2008). However, these suggestions are ineffective without informed exploratory research that analyses the situation and clearly diagnoses the problem, followed by recommendations for the appropriate interventions. This literature review indicates that the OSCE is a growing field of study in SA, and further research in this area has the potential to significantly impact clinical assessment in the country (Al-Rubaish et al., 2005; Alaidarous et al., 2016; Alsaid and Al-Sheikh, 2017).

Challenges

According to the findings of this scoping review, new medical schools seem to struggle more with OSCE implementation than well-established schools with a significant amount of experience. Some new medical schools host a large number of new staff who design and run their OSCEs. Those staff may not receive adequate faculty training on developing quality assessments. Furthermore, new medical schools may lack experience, regulations, funds, and infrastructure. These factors may explain the variation in OSCE implementation among medical schools (Alsaid and Al-Sheikh, 2017; Badawi et al., 2023). An additional issue is the global lack of clear guidelines or well-justified recommendations on conducting high-quality OSCEs (Khan et al., 2013b).

Due to the complexity of designing and implementing OSCEs, long preparation periods of at least 2–6 months are needed (Harden and Gleeson, 1979; Abdulghani et al., 2015); reasonable resources, experience, personnel, and support are required, which may be possible in some educational cultures, schools, and hospitals. However, some Saudi medical schools lack the requisite support and expertise. Due to a lack of experience, the culture in some medical schools may dictate that 2-3 weeks or even a few days is adequate preparation time (Elfaki et al., 2008), resulting in a suboptimal preparation phase (Ansari et al., 2021). They may also experience a deficiency in proper central coordination and a regimented manner of preparation that can ensure all OSCE requirements work in synchronisation, for example, advance notice for busy clinicians, recruiting SPs, and allocating equipment. Therefore, the feasibility issue related to the OSCE is still a challenge hindering its development (Al-Naami, 2008).

Further exploration of clinical assessments revealed that some medical schools in SA use only one clinical assessment tool, which is OSCE in some cases (Moeen-uz-zafar et al., 2015). However, no tool can measure all aspects of any domain, such as clinical competency (van der Vleuten, 1996; Harden, 2015). For the sake of ensuring that medical students have been assessed thoroughly and fairly, medical schools should implement integral assessment instruments to sufficiently evaluate clinical competencies (Khan et al., 2013b).

Exploring the OSCE literature in SA reveals that there is often little evidence and guidance to work with, and there is no evidence that experts in OSCE or assessment are available in every medical school. This scoping review revealed no evidence of cooperation between medical schools to share expertise or implement a feasible and valid OSCE. This may be due to a lack of awareness among faculties regarding the significance of assessment specialists or the cost of accommodating these experts, hindering faculties from improving their assessments in general and the OSCE in particular. These challenges call for an in-depth look at the OSCE in the Saudi context.

2.3.2.3 Summary of phase two

Patrício et al. (2009) complain about the low quality of the papers published on OSCE internationally, as some important information is missing, and there is inconsistency in reporting. I observed these issues to a greater extent in numerous OSCE-related papers in Saudi literature. Additionally, most papers are descriptive, and there is insufficient discussion or explanation of the OSCE design and implementation process, possibly because of the journal's limitations and focus. However, apart from some descriptions of the OSCE implementation in some papers, most papers discuss the students' opinions and conduct statistical analysis on the exam results, so there is no comprehensive understanding of the status of the OSCE implementation in SA that can be drawn. Furthermore, most papers lack information that would allow readers to comprehend the implementation process, background information, and how these factors contributed to the study results. Therefore, the picture of the OSCE implementation is not very clear, and drawing conclusions about this issue merely from students' perceptions and exam results will be difficult and inaccurate.

In reviewing the studies described above, I have found that OSCE implementation in SA has not been appropriately investigated; all the available papers illustrate individual medical schools' experiences with their OSCEs, but they tend not to follow clear guidelines or evidence-based practice, nor do they provide clear details and justifications for their choices during the design and implementation process. Implementation issues like blueprinting, station writing, designing scoring rubrics, and deciding standard settings are not sufficiently discussed. Therefore, no clear evidence is provided as to

whether medical schools in SA have implemented high- or low-quality OSCEs; thus, the OSCE quality may be compromised.

More than half of the studies (18 out of 30 included studies) were published in non-indexed journals, according to the National Library of Medicine (NLM) Catalogue (PubMed, 2023), which may indicate the low quality and rigour of these papers. Furthermore, the effect of different elements of the OSCE implementation has not been examined collectively in any one study in the Saudi context. This highlights the significant need for such empirical evidence.

After reviewing many papers on OSCE validity from respected researchers across many countries, Hodges (2003b) discovers that the contextual effects of economic, political, social, and cultural factors potentially influence OSCE validity. This researcher calls for context-specific and sophisticated qualitative investigations to examine the OSCE quality in a given context. The need for exploratory research to evaluate performance assessment, namely the OSCE, in the Saudi context is clear.

2.4 Summary of the Literature Review

From this inductive scoping review, I found that there are a few good-quality publications about the OSCE, most of which are from Western settings, and some are better than others. However, it seems there is no single, unified, comprehensive guide to the OSCE that takes the context of Saudi medical schools into account and enables them to implement the OSCE with high quality. Most OSCE-related resources can be viewed as a patchwork of OSCE people, processes, and papers, which schools must access and triangulate to implement the OSCE. The recent implementation experience of some countries highlights the opportunities and challenges that schools newly adopting the OSCE may face. This review shows that there is a need to explore why different schools in SA conduct OSCEs differently, understand what distinguishes effective OSCEs from low-quality OSCEs, and why. Such findings would be invaluable because they would support the use of evidence-based practices by defining criteria for effective methods of implementing the OSCE in SA. Although OSCE applications are discussed reasonably thoroughly in Western settings, there is a need for further discussion and improvements in SA due to cultural and geographical differences and because there is a small base of literature that discusses the design and implementation of OSCEs in the Saudi context.

2.5 Research aim, questions, and objectives

My personal experience of Saudi medical schools and my exploration of the literature have meant that I have been able to define the following research aim, questions, and objectives, which are presented below.

2.5.1 Research aim

This research aims to explore the implementation of the OSCE in the Saudi context and develop recommendations for improvement, which medical schools can use to improve their OSCE implementation and produce more defensible results.

2.5.2 Research questions

The research questions of this study are:

- How is the OSCE being implemented in two Saudi medical schools?
- What are the opportunities and challenges offered by adopting the OSCE in these medical schools?

2.5.3 Research objectives

The research objectives of this study are:

- To analyse how some Saudi medical schools design and conduct OSCEs using assessments and guidelines described in the literature.
- To identify how these Saudi medical schools ensure OSCE quality.
- To explore the opportunities and challenges of designing and implementing the OSCE in the Saudi context.
- To contribute to the body of knowledge on OSCE implementation, share study findings with research participants, and inform other medical schools with this study suggestions and recommendations.

2.5.4 Developments in the study's questions and objectives

The original research questions and objectives have evolved over the course of this study. They were originally stated in a way that implies that the findings and results are generalisable. For example, the original research question was: How is the OSCE being implemented in Saudi medical schools? However, my understanding has evolved, and the philosophical paradigm that I have adopted cautions against making generalisable statements or providing generalisable findings without taking the specific context into account. The constructivist paradigm (see Section 3.2.1) recognises that what works in one context may not necessarily work the same way in another (Cohen et al., 2018). As a result, this paradigm deals with the data and findings by using transferability rather than generalisability (Lincoln and Guba, 1985; Varpio et al., 2021). Therefore, I have clarified the research questions and objectives and identified the specific contexts that have been investigated.

Chapter 3 Methodology and Methods

3.1 Introduction

In this chapter, I set out the methodological approach used to address the research aim. I describe the philosophical stance that underpins the methodology I adopt and inform the selection of methods and data analysis approaches to best serve the research aim, questions, and objectives. I describe the qualitative research design and case study approach that I employed for this project. I then discuss sampling, recruitment, ethical approval, pilot testing, and data collection procedures. Following this is a description of how I conducted the documentary, interviews, and focus groups (FGs) analyses and how I adhered to the quality criteria of qualitative research throughout the study.

3.2 Research Methodology

This section outlines the research philosophy, design, and method selection.

3.2.1 Research philosophy

In this empirically-based research project, I investigate the quality of the OSCE in Saudi medical schools by examining the quality of its implementation. By understanding how OSCEs are designed, organised, and delivered, I am able to identify their strengths and weaknesses. This is what Pell et al. (2010) call ‘assessing the assessment’, which is crucial to ensuring quality given the high-stakes nature of medicine and its summative clinical assessments. There are several philosophical stances on which the research could be based, including positivist, post-positivist, and constructivist paradigms (Tavakol and Sandars, 2014). Although both positivists and post-positivists believe in one truth in the social realm, the latter believe it cannot be truly observed. However, constructivists believe in multiple truths, which largely depend on context and how people interact with things (Bergman et al., 2012), which appears more in line with this study's research questions and objectives.

Therefore, the research paradigm underpinning my investigation is the constructivist (also known as interpretivist or naturalist) theoretical stance (Creswell, 2014; Cohen et al., 2018). The ontological (theory of reality) assumption of the interpretivist paradigm is that the reality of something exists, but it can be multiple and is constantly changing (Cohen et al., 2018). The epistemological (theory of knowledge) nature of constructivist is that the ideal knowledge is subjective and can be generated through direct interaction between the researcher and participants (Cohen et al., 2018). Because objectivity cannot be achieved, an interpretive and constructive approach should be used to approach the investigation of reality (Bergman et al., 2012). Therefore, qualitative approaches can

generate ideal data to address the research issue (Lincoln and Guba, 1985; Bergman et al., 2012).

The constructivist paradigm justifies the rationale behind using the qualitative method approach and explains how the approach suits research issues similar to mine (Denzin and Lincoln, 2018). The qualitative approach is more likely to be able to investigate OSCE implementation in medical schools because there is no predetermined hypothesis about the situation to be examined (Patton, 2015). The qualitative approach supports the use of an inductive style to generate knowledge and conception from the data rather than testing preset hypotheses deductively (Cleland and Durning, 2022). Having a clear conceptual framework not only informs the research methodology but also underpins the data analysis framework (Varpio et al., 2020). One dataset collected by one method cannot clearly explain this issue, but triangulated data collected by more than one instrument can help to understand reality better (Blaikie and Priest, 2019). Using multiple data collection instruments and interpreting data from multiple sources will strengthen research findings' reliability, internal validity, and trustworthiness (Creswell, 2014). This approach gives the study the necessary depth to enrich its findings, provide opportunities to test a model or theory and increase confidence in the study's findings (Hanson et al., 2005).

3.2.2 Research design

Since the research questions seek to investigate the OSCE implementation in Saudi medical schools, it is necessary to visit different medical schools to investigate this topic, as the experiences and contexts of the schools differ. Therefore, I need to collect similar data from each school. According to Yin (2018) and Schwandt and Gates (2018), the case study approach is a suitable choice in such a situation, with the medical school acting as the case. The more case studies included, the better it would be; however, due to feasibility issues, such as being the sole researcher with limited accessibility, resources, and time, I am limited in the number of case studies I can produce and the type of data that I can collect. For instance, more than two case studies would disperse the researcher's focus and diminish the richness and depth of the qualitative analysis. Furthermore, most doctoral educational research focuses more on depth than breadth when studying a phenomenon under investigation (Kirchherr, 2019). Additionally, the study paradigm and aim guide the study to conduct a qualitative but not quantitative approach and investigate the research question in depth rather than breadth to provide generalisable results (Flick, 2019). Hence, I conducted two case studies using multiple qualitative data collection instruments.

3.2.3 Research methods

To ensure alignment between my methodology and methods, I decided to use the data collection instruments underpinned by the constructivist research paradigm (Creswell, 2014). Within the qualitative approach, there are several methods or instruments for data collection, including interviews, FGs, documentary analysis, ethnography, and visual methods (e.g., drawings, photos, and videos).

I examined the viability of each method and determined that visual methods may be better suited for research that seeks the perceptions and experiences of individuals, but they also lack contextual information, which would not be suitable for my specific research interest (Cristancho et al., 2022). The ethnographic approach, on the other hand, can provide in-depth context understanding and a holistic perspective for the investigated phenomenon; however, it requires extended periods of fieldwork from the researcher, which would not be suitable for a single researcher with a limited timeframe and the intent to conduct multiple case studies (Kitto et al., 2022). Both interviews and FGs allow for in-depth exploration, are flexible in structuring and questioning, enable the researcher to interact with multiple participants, and generate rich and triangulated data (Cohen et al., 2018; Denzin and Lincoln, 2018; Flick, 2019). Documentary analysis is also a potent source of data for assessment research, as it can provide the research's background by assessing the existing data, identifying gaps and discrepancies, and triangulating the data with other sources (Gorsky and Mold, 2020). Therefore, I decided that each case study would involve documentary analysis, interviews, and FGs (Figure 8). Triangulation between methods used for data collection would provide information from different sources, enhance the trustworthiness of the collected data, and establish the credibility of the findings (Stalmeijer et al., 2014).

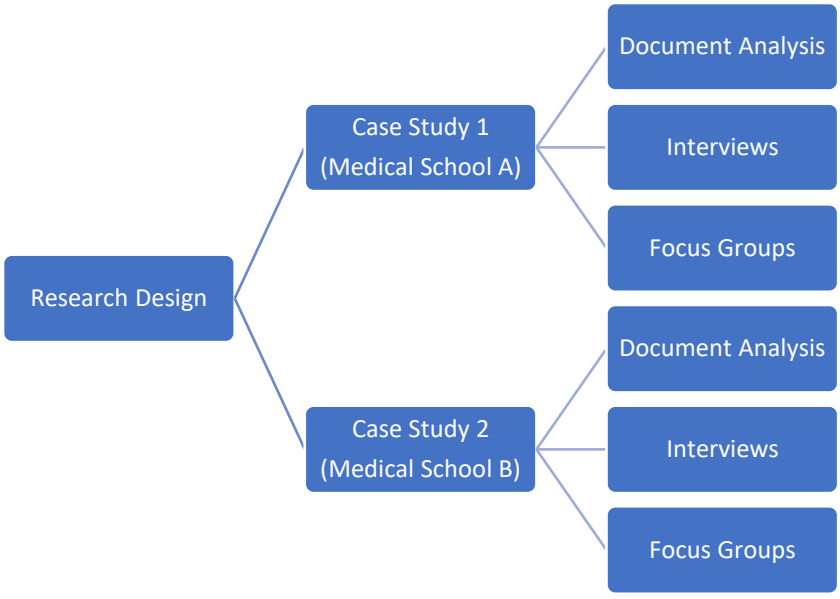


Figure 8. Research design and methods outline.

3.3 Case studies design

The case study approach provides a richer and clearer triangulated view of the research issue and will produce stronger evidence than using a single case study and a single instrument (Yazan, 2015; Yin, 2018). I find this approach useful because it allows me to treat each medical school (case study) as an analysis unit, considering its context, unique features, and local OSCE practice (Stake, 2006). Yin (2018) suggests a case study is appropriate when: a) a unique context is established that produces a particular phenomenon; b) multiple data sources are available for data collection; and c) multiple data collection methods are employed. In this study, medical schools are the case, OSCE implementation is the phenomenon, assessment documents and OSCE personnel discussions are the data sources, and document analysis, interviews, and FGs are the data collection techniques.

Therefore, I conducted two exploratory case studies, applying multiple qualitative data collection methods to each. I first collected, analysed, and interpreted data from each case study separately (Cohen et al., 2018). Then, I utilised the exploratory case study analytic strategy described by Creswell (2014) and Yin (2018) by engaging with the data, generating the findings, and comparing data from different datasets within the same case study to determine whether the results support or refute each other. Lastly, Yin (2018) and Stake (2006) suggest using the cross-case synthesis model as an analytic technique so that results from both case studies can be combined, compared, and discussed to draw comprehensive results and conclusions (Figure 9).

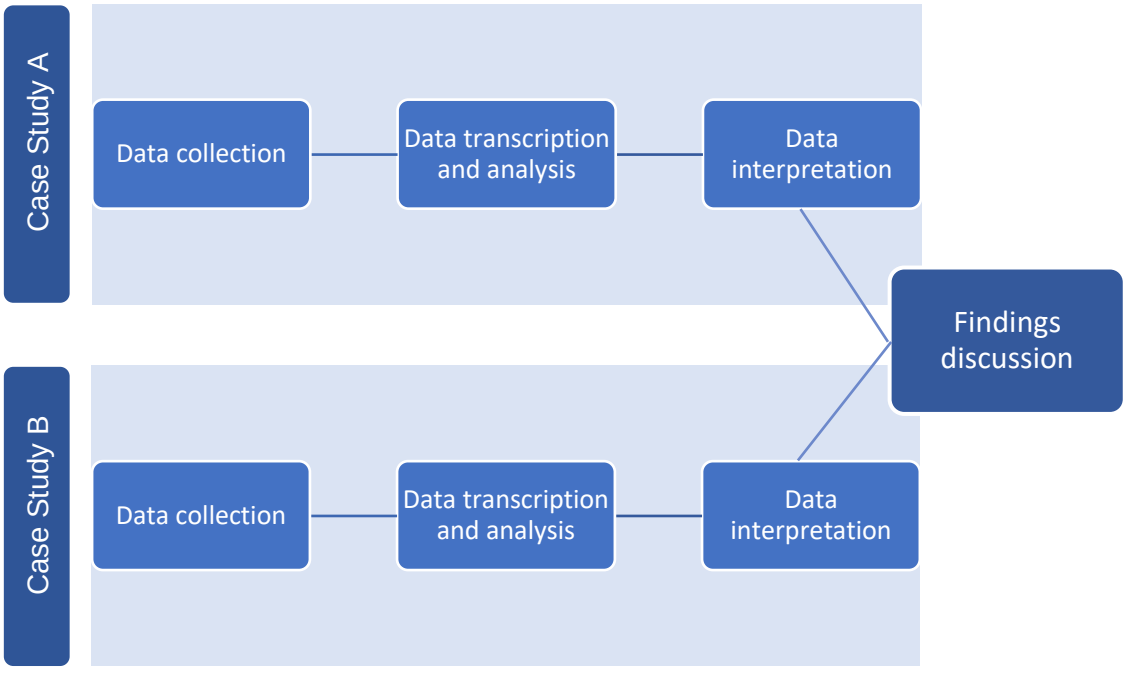


Figure 9. The analytic strategy.

I examined the two schools' assessment documents and the individual and group experiences of various stakeholders to obtain diverse perspectives on OSCE implementation. This provided copious data that allowed me to thoroughly comprehend the research issue. As a result, I could analyse, compare, and contrast these perspectives and draw meaningful conclusions to evaluate the OSCE implementation at participant schools.

3.3.1 Selecting the case studies

Case selection is vital for the study; I considered this and provided a rationale for the selection criteria for participating medical schools. This study's focus is on medical schools in SA that use the OSCE as a summative assessment for their medical students. Collecting data from two medical schools in different contexts would provide valuable insight into this study's focus (Creswell, 2014; Flick, 2014). Different models would broaden the range of data sources available to aid in answering the research question. In light of this, I looked at the medical schools that exhibit variations in key characteristics. Differences in location (choosing schools from different cities), resources (public versus private schools), and accreditation statuses (accredited versus not accredited) are factors that can enrich case study findings. Geographical locations may influence how the OSCE is implemented due to contextual and relative cultural differences between different locations (Hodges, 2003b). Data from public and private medical schools could clarify the differences in how the two systems implement and administer the OSCE (Hernandez-Galvez and Roldan-Valadez, 2019). Likewise, NCAAA accreditation may impact how the OSCE is implemented and how schools ensure assessment quality (Alrebish, 2014). Nevertheless, constructivist approaches aim for conceptual understanding and case-by-case transferability rather than statistical generalizability (Cohen et al., 2018; Varpio et al., 2021).

3.3.1.1 Context of the case studies

Using the aforementioned criteria while still considering accessibility and feasibility issues, the first medical school to which I was able to gain access is a public, non-accredited medical school in a medium-sized city. The second case involved a private, accredited medical school in a smaller city. Here, I provide an overview of each medical school's context while avoiding providing specific or detailed information that could reveal the school's identity (see the ethical approval section 3.5).

Case study A

The main characteristics of case study A are that it is a newly established medical school that was established about ten years ago. It is exclusively funded by the government and must therefore adhere to the standards and regulations of the Ministry of Education. This medical school has a capacity of just under one hundred students per cohort, with a

male-to-female ratio of one to one, which may make it a small medical school in Saudi terms. This school adopted a five-year bachelor's programme (three years of basic sciences and two years of clinical sciences) and a student-centred education approach. However, the school has no accreditation and no partnerships with other medical schools nationally or internationally. The OSCE is administered as the only summative performance assessment at the end of each clinical course, with between two and five stations per course.

Despite the presence of senior and experienced medical staff, the vast majority of teaching and administrative staff are young local faculty clinicians. However, the majority of them are newly appointed with limited to no prior experience in higher education teaching or leadership. This general context is shared by roughly 31 public medical schools of SA's 41 medical schools (Sebai, 2019). As a result, this medical school may have many similarities with 26 medical schools in SA (31 new public schools minus 5 old schools); thus, I considered this to be the most common type of medical school in SA.

Case study B

The context of case study B differs from that of case study A in several ways. Although case study B's medical school is also new (about ten years old), it is a private, non-profit institution. It is located in a different but smaller city and is considered a small to medium-sized medical school because each cohort has less than 150 students. Most of the faculty are senior and international staff who have gained various experiences before joining this medical school. As in case study A, this medical school has a five-year bachelor's programme (three years of basic sciences and two years of clinical sciences), and a student-centred approach to education. All courses in this medical school use the OSCE as the sole summative performance assessment, with five to six stations per course.

This medical school has full local academic accreditation (from the NCAAA) as well as international accreditation. These accreditations are meant to reflect the medical school's quality assurance measures and high standards in teaching and administration. It also has a partnership agreement with a well-known Western medical school, which may benefit the school's curriculum, teaching, and assessments. This suggests that medical school B worked to match the standards of this international school; however, to protect the anonymity of these institutions, I chose not to share their names because doing so could reveal the identity of this medical school.

Case study B is an example of, and shares the context with, the growing type of medical schools in SA, which is estimated to be ten medical schools nationally (Sebai, 2019). This type of medical school has many similarities, which can be summarised in two areas: a) private medical school with an independent fund; and b) adaptable and flexible

in introducing educational or administrative changes as it is not required to follow the government's administrative structure.

3.4 Methods and sampling

This section outlines the rationale for using documentary analysis, interviews, and FGs as data sources, as well as the sampling and recruitment processes.

3.4.1 Documentary analysis

Document analysis is often used to triangulate findings gathered from other data sources, which in this research are the interviews and the FGs (O′Leary, 2021). Bowen (2009) argued that document analysis can widen the coverage of an investigation on a certain topic by providing information that might not be gathered by other means, e.g., interviews and FGs. It is an effective way to contextualise the research and ensure its rigour and comprehensiveness (Freeman and Maybin, 2011; Walsh, 2014). When utilised in triangulation, documents can validate, clarify, or refute the findings of other data sources (Bowen, 2009). On the other hand, document analyses cannot always answer the research question because the documents under analysis were not created for my research; consequently, I may not find exactly what I need in my research. However, it is useful to outline the available and missing information in these documents.

I requested assessment documents used to design and implement the OSCE at the participating medical schools. (Silverman, 2017). The documents that I gathered to help me understand how schools implement the OSCE include the following:

- School assessment policies and regulations,
- OSCE design information and templates,
- OSCE blueprints, scenarios, and instructions,
- OSCE marking schemes,
- And other documents that help in understanding the school's OSCE implementation.

A detailed list of each case study's documents is provided at the beginning of each case study analysis chapter (Chapter 4 and Chapter 5). I gathered as many assessment components as possible from each medical school. I received authorisation from the gatekeepers, i.e., the deans of medical schools (refer to Section 3.5), so they forwarded my request to the faculty to provide the information I requested. I collected documents from a variety of personnel, including the head of the assessment unit, the head of the Health Professions Education (HPE) unit, and OSCE designers. Because I received the documents from various sources, there were many repetitions; thus, I initially filtered out the repetitions and unrelated documents.

Conducting a documental analysis (explained in detail in Section 3.8.1) of these documents allowed me to compare mere rhetoric to the participants' perceptions. For example, documents emphasise that assessors should perform a thorough blueprinting for the OSCE, but after triangulating that with FGs and interviews findings, it turns out that they do not perform a proper blueprinting (4.3.2.2). I analysed these documents to understand the context, learn how the medical schools implement and deliver their OSCEs, and assess content validity to ensure the OSCEs assess the appropriate skills and knowledge (Schuwirth and Vleuten, 2011). For example, to determine whether OSCE blueprints, scenarios, and scoring schemes align with the criteria for a good assessment design outlined in the literature review (Harden et al., 2015). The documentary analysis explored quality measures the schools use in their quality toolkits, for example, which psychometric analysis they use to ensure exam reliability and how they determine standard setting (Pell et al., 2010). In addition, I reviewed the clarity of assessment implementation policies and regulations, the instructions for examiners and SPs training, and the conformity with the criteria of a good OSCE as outlined in the literature (Khan et al., 2013a). I reviewed the documents before conducting the interviews and FGs, and I also analysed the documents before analysing the interviews and FGs. The findings of the document analysis provided an essential overview of the current OSCE implementation at each medical school and assisted me in comprehending the situation during the analysis of the data from the interviews and FGs.

3.4.2 Interviews

Semi-structured interviews have largely enriched the research and enabled me to ask open-ended questions to elicit deep insight regarding different aspects of OSCE implementation (Silverman, 2017; Cohen et al., 2018). While both structured and semi-structured interviews can facilitate in-depth exploration, structured interviews may generate irrelevant data because the interviewer cannot intervene during the interview, whereas the semi-structured interview format allows the interviewer to guide the conversation to benefit the research (Kvale, 2018). The semi-structured interview provides the flexibility needed in this research to emphasise some points, ask follow-up questions on others, and allow the interviewer and interviewee to check each other’s understanding, enhancing the collected data's content validity (King et al., 2018). The interview is a time-consuming method, requiring a similar time commitment to FGs but only eliciting data from one participant. However, it provides more in-depth insight, is easier to arrange, and can be conducted remotely under certain circumstances (Rubin and Rubin, 2012).

After reviewing the document analyses and gaining a basic understanding of what the document offers regarding OSCE implementation, I proceeded to gain in-depth insight into the actual OSCE implementation by conducting semi-structured interviews with

other stakeholders (i.e., the educational administrator, the head of assessment and medical education unit, and local assessment specialists) (Flick, 2014). FGs are unsuitable for staff at their hierarchical level; each one faces different challenges and holds the power to influence OSCE implementation (Rubin and Rubin, 2012). One interview in each school was with the person responsible for exams and assessment, e.g., the assessment unit and medical education department leaders. In addition to the opinions of OSCE designers and examiners (from the FGs), these individuals provided me with a different viewpoint on how OSCEs are designed and implemented in their schools. The assumption that everyone in a school works in harmony may not be true in all contexts (Lee, 2016). Hearing from different people at various levels of authority may result in a clash of perspectives, clarify ambiguity, confirm views, or evoke further investigation.

I conducted another interview in each school with a medical school's administrative leader, such as the dean or vice dean for academic affairs. These interviews informed the research about the rigours and requirements of the higher administrative hierarchy to determine whether the structure supports or inhibits the autonomy of the school to make its own decisions in implementing the OSCE. In addition, they contributed to the research on the impact of medical school administration on the OSCE and their ability to support OSCE improvement.

Local assessment specialists within each school can provide valuable insight into current practises. I conducted these interviews last, after completing the other interviews, as I had acquired sufficient knowledge of the OSCE practises in each school's context from a variety of participants and sources. These individuals can then provide clarifications and responses to questions that other participants cannot address (Cohen et al., 2018). They were able to explain the unwritten rules and procedures they use to manage the OSCE and provide justification or scientific evidence where available. Furthermore, they provided suggestions and actionable steps to improve their OSCE implementation. Figure 10 summarises the interviewed groups.

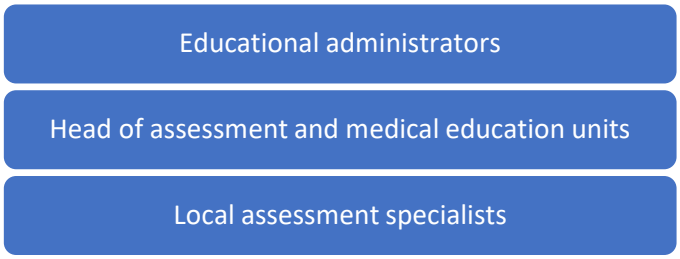


Figure 10. Interviews participants.

3.4.3 Focus groups

In the OSCE literature, the staff member responsible for identifying the purpose of a particular OSCE, preparing a blueprint, writing questions and instructions, and documenting requirements for the OSCE is referred to by different titles, including designer, writer, organiser, and planner (Boursicot and Roberts, 2005; Gupta et al., 2010; Khan et al., 2013a; Harden et al., 2015). According to the Merriam-Webster (2019) dictionary, a designer is the 'one who creates and often executes plans for a project or structure'. Thus, in this research, I will use OSCE *designer* to refer to persons who are designing, writing, organising, planning, and preparing OSCEs; I will use *examiner* to refer to those who observe students' skills and score their performance.

FG engages multiple participants with a shared interest in a group discussion on a particular topic (Marshall and Rossman, 2014; King et al., 2018). FGs are useful in providing collective views through interactive discussion, as opposed to the individual opinions obtained when subjects interact only with the interviewer (Rubin and Rubin, 2012). FG facilitates group dynamics to share experiences and engage in meaningful discussion (Cohen et al., 2018). Although FGs can be difficult to organise, they seem more time efficient and produce more data from multiple perspectives simultaneously compared to face-to-face interviews (Barbour, 2008). However, both FGs and interviews permit validation and clarification to reduce misunderstandings and increase the information's credibility (Marshall and Rossman, 2014).

I looked for the main OSCE personnel in medical schools and found them to be the OSCE organisers (including station writers) and the examiners. However, their roles in the implementation process are distinct; thus, I conducted a separate case study for each group. With respect to Saudi Arabian cultural values, it is more appropriate and convenient for the participants to conduct separate FGs for male and female participants, which I felt allowed each group to speak freely and candidly. Thus, I conducted two FGs for the OSCE organisers and two FGs for the OSCE examiners. I employed the same methodology for both case studies, resulting in eight FGs in total across both medical schools (Figure 11).

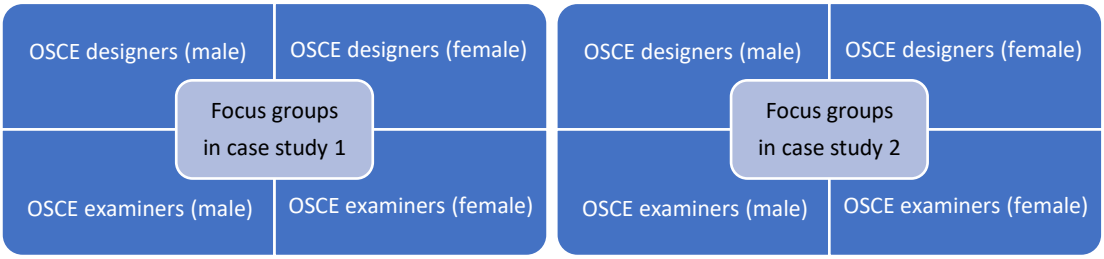


Figure 11. Focus groups sampling.

Because the population of interest in this research scenario is small, I invited five to seven participants to each FG, as fewer or more participants could have an unfavourable impact (Bryman, 2016). For example, two or three participants may not initiate a favourable group dynamic and might yield the FG to an interview. On the other hand, more than eight participants could be impractical and difficult to manage (Cohen et al., 2018); however, the actual participation ranged between four and six.

During the FG discussion, I acted only as a facilitator to steer the discussion but did not participate in the discussion unless necessary, for example, when I needed to clarify an issue or refocus the group on the research interest (Cohen et al., 2018). I also tried to maintain a fruitful group discussion, control dominant participants, encourage quiet ones, and allow enough time for participants to clarify their thoughts (Krueger and Casey, 2015). I followed Bryman's (2016) recommendation of taking field notes on group dynamics, capturing agreements and disagreements, documenting verbal and non-verbal languages, and gathering useful information during data interpretation.

I designed three question guides in total: one for designers' and writers' FGs (see in Appendix 10 for an example), one for examiners' FGs, and one for interviews, which are all based on the literature review findings (Marshall and Rossman, 2014). Using carefully crafted question guides improves the consistency of the collected data and ensures a balanced discussion of important aspects of the research, thereby enhancing data validity and reliability and reducing researcher error (Denzin and Lincoln, 2018). All question guides are organised into clusters of questions under the following three headings: opening and introductory, key content exploration, and closing questions (Bryman, 2016). The warming-up questions enabled me to gain participants' trust and stimulate them to speak freely and candidly (King et al., 2018). Allan and Skinner (1991) suggest starting the FGs or interviews with general questions and leaving complicated or relatively sensitive questions for later in the discussion when rapport has been established. The exploratory and key questions, which were generated from pertinent literature and are relevant to research objectives, ensure content validity and serve the core interest of this research (Cohen et al., 2018). The closing questions gave the participants the chance to clarify, emphasise, or provide additional comments to wrap up the discussion, as well as ensure that no important information was missed (Bell and Waters, 2014). Figure 12 is an overview of the data collection methods, sources, and numbers.

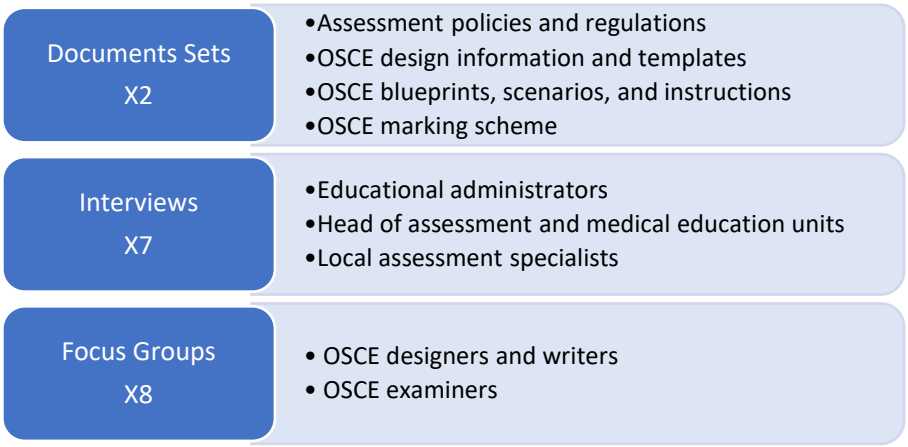


Figure 12. Overview of data collection methods, sources, and numbers.

3.4.4 Sampling and recruitment of the interviews and FGs participants

Choosing appropriate research participants is critical to ensuring research credibility (Morse et al., 2002). The research question and the study design determine the relative sample size so that only a specific group of individuals can inform the research purpose (Creswell, 2014; Cohen et al., 2018). Because only particular people can inform the research, I need to select them purposefully (Leavy, 2017). Therefore, participants will be chosen using the non-probability purposive sampling technique for both interviews and FGs (Blaikie and Priest, 2019). Although I estimated the approximate sample size for the interviews and FGs based on the research question and the number of people who are informative to the study, the final decision was made late in the data collection process. The nature of the constructivist qualitative research, the context of the medical schools, and my review of comparable research indicated that the sample size I obtained was sufficient (Boddy, 2016). Additionally, I benefited from the ground theory principle of data saturation in deciding the sample size because, after data collection, I believed that later meetings would not provide any additional important information, so I determined that the sample size was sufficient (Denzin and Lincoln, 2018). **Table 1**Table 4 show the sample size for interviews and FGs.

I interviewed educational leaders, the heads of the medical education departments and assessment units, and assessment specialists. I identified the assessment specialists based on the recommendations of others within each institution, as this individual could be regarded as knowledgeable in OSCE at that school (Cohen et al., 2018). For the FGs, I conducted multiple FGs in each school, two for OSCE designers (one for males and one for females) and two for OSCE examiners (one for males and one for females) due to the gender-separation culture in the school environment. As described in Section

3.4.3, the OSCE designers are those who plan, design, and oversee the entire OSCE process; the station writers are those who write the station scenarios and participant instructions; and the examiners are those who sit in the station to observe and assess students' performance. I requested from the head of the medical education units that they send invitations via email to the individuals I had identified as my target group. Then they responded with the names of those who accepted the invitation, after which I coordinated with the participants the meeting's time and method (see Section 3.7 for more details).

Table 4. This table shows the sample size.

	Participants	Case study A	Case study B	Total
Interviews	Academic leaders	2	1	3
	Assessment specialists	2	2	4
FGs	OSCE designers	2	2	4
	OSCE examiners	2	2	4
Total		8	7	15

3.5 Ethical issues and approval

The University of Leeds granted ethical approval for this study (Appendix 11). Consequently, I emailed a formal request to the deans of the two targeted medical schools (Appendix 12), as they are the gatekeepers in this situation, to obtain permission to conduct this study in their schools (Cohen et al., 2018). They both granted me approval to conduct the research at their schools and put me in direct contact with the heads of the medical education units, who then put me in touch with the participants. I sent all participants an invitation email containing copies of the information sheets and consent forms, which they all read, signed, and returned. The information sheets (Appendix 13) and consent forms (Appendix 14) contain all necessary information, such as the purpose and procedure of the study, the participants' responsibilities, rights, and benefits, and how I will handle their data (Bell and Waters, 2014). Despite the difficulty of ensuring the anonymity of individuals and schools due to the small sample size, I have guaranteed confidentiality to both medical schools and individual participants, emphasised that participation is entirely voluntary, and anonymised all data (Silverman, 2017). As a result, I am cautious when describing the medical schools that represent the case studies, as certain details could compromise their anonymity. Therefore, I took some precautions to avoid this, such as using pseudonyms for each medical school (i.e., medical schools A and B) and not disclosing the exact establishment date, student numbers, or the name of the teaching approach. These characteristics could be tied together to precisely identify the medical school.

To ensure participants' privacy, anonymity, and safety, I assigned pseudonyms to each interviewee, and they had complete autonomy to complete or withdraw from the study without providing a reason or incurring consequences (Bryman, 2016). I confirmed to participants that there would be no right or wrong answers and that their participation would not harm or exploit them in any way. I also assured them that if the research is published, it will not contain information that identifies them or their school (Cohen et al., 2018). For instance, I did not reveal their exact position or speciality because doing so would reveal their identity. I believe that the aforementioned measures alleviate participant worries and allow them to communicate their thoughts openly, thereby enhancing the validity of the data and addressing bias concerns (Flick, 2014). Furthermore, I adhered to the University of Leeds regulations and maintained a high level of information safeguarding by storing all documents, recordings, and other electronic forms of data in an encrypted format on the University of Leeds OneDrive™ cloud service.

3.6 Pilot testing

I conducted a pilot interview, which improved my interviewing abilities and also assisted me in leading the FGs (Flick, 2014). I conducted this pilot interview with a colleague who is familiar with the culture where this study will take place, and then I asked him for detailed feedback on the interview process. This feedback helped me enhance my interview skills, and I believe it increased the validity of the data I collected (Marshall and Rossman, 2014). The data collected in interviews are frequently criticised due to interviewers' biases and interviewers' communications, such as body language; the pilot study's feedback raised my awareness of these issues and assisted me in mitigating or even eliminating them (Kvale, 2018). Additionally, the piloting process allows me to evaluate the suitability of the question guides, allowing me to reword, add, or remove questions to ensure that the guide serves its intended purpose. This pilot was conducted online via Zoom™ to simulate the actual interviews, as described below (3.7), and to familiarise myself with this platform, which allowed me to examine the platform and examine the technical issues, such as the clarity of voice, videos, and recording system. This also helped me in conduction the FGs, as they are both types of interviews with similar interview protocols (Marshall and Rossman, 2014).

3.7 Conducting the interviews and FGs

The COVID-19 pandemic forced many countries into complete lockdown, including SA and the UK. This unprecedented situation created a high level of uncertainty at all levels. For instance, the regulations of the University of Leeds prevented travel for research purposes during this pandemic. Additionally, all educational institutions were closed during the lockdown. This situation made online conferences, meetings, and interviews

the new norm, as it maintained the social distancing regulations. As a result of COVID-19's uncertainty and constraints, conducting interviews and FGs online became the only viable option. My supervisors and I discussed this matter and agreed that online data collection was necessary to avoid further delays.

Then, I started contacting the targeted group I had already identified, as explained earlier (3.4.4). I contacted them via email and allowed the participants to choose the most convenient time for them. I found Bryman's (2016) note about the challenges of negotiating multiple participants to set for the FG at the same time to be accurate; in contrast, allocating time with interview participants was straightforward. However, after exchanging multiple emails with each FG member, I was able to allocate the time that was best for each group.

I have explored several platforms for teleconferences, including Microsoft Teams™, GoToMeeting™, Cisco WebEx™, and Zoom™. I found that Zoom™ is the most suitable platform for several reasons, including:

- It provides a stable and clear connection and excellent audio and video quality.
- It is easy to use and compliant with the University of Leeds's encryption requirements (end-to-end encryption).
- It provides built-in recording functionality that enables me to record the entire conversation clearly. Zoom™ recording is superior to the actual conversation because it records the speaker's side and combines the conversation at the end of the recording. Therefore, there is no over-the-internet recording, which could experience sudden voice loss due to a poor online connection.
- Everyone I met was familiar with it and ready to use it.

Consequently, all interviews were conducted and recorded using the encrypted version of Zoom. I used the recording functionality for the entire conversation as well as recording the conversation from my side using an audio recorder to avoid any unforeseen technical errors (Bryman, 2016). I gave a unique pseudonym for each record and stored them all in the University of Leeds OneDrive™ cloud service.

To a large extent, I felt that the online meeting was similar to the face-to-face meetings. Although there are some variations, I do not believe I have lost any significant advantages by conducting online meetings. I purposefully engaged in informal conversations with the participants prior to the interviews to warm them up for the discussion and establish rapport (Patton, 2015). The online interviews and FGs went smoothly, and the sound and video quality were excellent. The interviewees' facial expressions were clear, and I felt they were open and cooperative. I believe the participants were at ease because they attended the meeting at their own convenient time and place (Cohen et al., 2018). Additional benefits of online meetings include time, cost efficiency, and environmental friendliness because they reduce the cost and time

required for travel and logistical arrangements to hold in-person meetings (Milic et al., 2020).

Although I have not experienced any significant downsides to online meetings, I have encountered a few minor ones. For example, an online connection was lost in the middle of one interview, but the participant quickly reconnected and continued with the interview without major disruption. A second issue was that a few participants on some FGs chose to close the cameras, so I was unable to observe their facial expressions and emotions. Nonetheless, after a gentle reminder, some individuals reopened their cameras and rejoined the productive discussion. I felt those who decided to close their cameras desired to be less visible and remain quiet. Therefore, when they remained silent for an extended period of time, I probed them and asked them to share their opinion in a gentle manner (Cohen et al., 2018), for example, Dr X, what do you believe to be the cause of the lack of examiner training? However, silent participants can also occur during face-to-face FGs, requiring researcher intervention (Bryman, 2016).

Most of the interviewees spoke Arabic or Urdu, and English was their second language. Although I gave the participants the option of speaking in Arabic or English, all interviews were conducted in English, apart from a few Arabic words or phrases. Consequently, I had no issues translating the transcripts. However, the varying English proficiency of the interviewees made some of the quotes difficult to understand, so I attempted to make minor adjustments with the help of the field notes to make the quotes easy to understand. I did so with the original meaning in mind, attempting to preserve the interviewee's original meaning. For example, I corrected the verb tenses and the singular and plural pronouns. I also changed a few words, such as 'external examiner' to 'outside examiners' (examiners from outside the institution who are not evaluators), to reflect their intended meaning and align them with the UK definition of external examiners. These minor changes were validated by a colleague from the University of Leeds who is very familiar with the context in which this case study takes place.

3.8 Data analysis

In this section, I describe the data analysis process and explain the methods I employed to analyse the two types of data I collected. For the assessment documents, I used the Codebook approach, and for the interview and FG data, I used the Reflexive Thematic Analysis (RTA) approach.

3.8.1 Documentary analysis

To utilise the assessment documents meaningfully informing the research focus, I decided to compare their alignment with the assessment literature (Shaw et al., 2004; Flick, 2014). However, the assessment literature is vast, making it impossible to compare schools' documents with the entire assessment literature. As a result, I chose to look for

well-known, widely accepted, and reputable organisations in medical education that have comprehensive assessment documents; rigorous standards; place special emphasis on assessment quality; and are constructed by experts who typically rely on evidence-based practises. Among several international medical education organisations, I found three that meet the aforementioned criteria and emphasise assessment implementation in medical schools. The three organisations whose criteria and frameworks I used are:

1- Ottawa Conference consensus framework for good assessment and the consensus statement and recommendations for performance assessment (Boursicot et al., 2011; Norcini et al., 2018; Boursicot et al., 2020).
2- Criteria of the ASPIRE-to-Excellence award in assessment of students by AMEE (AMEE, 2023).
3- The World Federation for Medical Education (WFME) global standards for quality improvement in assessment (WFME, 2015; WFME, 2020).

After identifying these institutions' assessment documents, I had to determine the most effective documentary analysis method. There is a spectrum of approaches to analysing these documents (Park et al., 2012; O′Leary, 2021). Therefore, I should make an informed decision that aligns with my purpose of analysing these documents, as each approach has advantages and disadvantages.

3.8.1.1 Deciding the analysis approach

I initially wanted to apply the same analytic approach to the interviews, FGs, and documents, as I thought this would ensure analytical consistency across the datasets. Therefore, I conducted an RTA on the assessment documents and wrote a report on the findings. The analysis results were useful; however, they did not necessarily inform the purpose of the assessment documents in understanding OSCE implementation in medical schools. Therefore, I began searching for an alternative method of analysis that could serve the purpose of this study. So I read about various methods, such as critical discourse analysis, which can potentially analyse documents (Bryman, 2016). I understand that critical discourse analysis tends to focus on verbal interactions and the role of language in social discourse and, to a lesser degree, formal documents (Flick, 2014). Bryman (2016) indicated that critical discourse analysis is most helpful when searching for different and latent meanings in narrative accounts; however, I was interested in the facts of what was mentioned and what was not mentioned about the OSCE. I discovered that critical discourse analysis may not be beneficial to my objective for document analysis. I want a methodology that specifically looks for what was stated in the OSCE assessment documents and can relate that to both the literature and other data gathered for this study.

Another possible approach to analysing the documents is coding reliability (Boyatzis, 1998). This method emphasises reliability to produce reliable codes, so it seeks

consensus from multiple researchers, as it is based on positivist principles, and it also generates codes relatively quickly and easily (Byrne, 2022). This indicates its incompatibility with my research paradigm and the need for multiple researchers to ensure inter-rater reliability, as well as its rigidity regarding what I can include and exclude, so it would include codes irrelevant to my research interests (Guest et al., 2012).

Further reading and consultation with methodology specialists at the University of Leeds were required to determine the best documentary analysis approach for the research focus of my thesis. This attempt led me to the codebook approach, which I thoroughly researched to determine if it was suitable for the purpose of this analysis. As the documents are not my primary source of data but rather a supplementary resource, the codebook approach proved useful in allowing me to be purposefully selective in what I looked for (Bowen, 2009; Roberts et al., 2019; Braun and Clarke, 2021). This strategy enables me, as a researcher, to prioritise and concentrate solely on OSCE- and regulations-related matters, as the assessment documents contained a great deal of information that is irrelevant to this study (Bowen, 2009; Mihas, 2019). Moreover, the codebook approach is consistent with the constructivist paradigm and has a structured coding strategy (Braun and Clarke, 2020). Therefore, I believe the codebook approach is appropriate because this research aims to understand the assessment system and how it relates to the OSCE as an assessment method rather than an in-depth investigation of all the documents' content.

3.8.1.2 Codebook analysis approach

Codebook analysis is a subset of thematic analysis approaches (King and Brooks, 2018; Roberts et al., 2019; Braun and Clarke, 2021). It is primarily characterised by a deductive analysis approach, in which the resultant codes and themes are typically inputs to the analysis processes; in contrast, RTA relies on an inductive analysis approach, so the themes are the outputs of the analysis (Crabtree and Miller, 1999; Mihas, 2019; Braun and Clarke, 2020). Consequently, in the codebook method, the researcher can identify all or most codes and themes early in the analysis. The themes produced using this method are typically more like domain summaries (hence I called them key areas, for example, ‘criteria of good assessment’ and ‘use a system of assessment’), which categorise the explicit content of the data rather than offering a richer interpretation of more latent meanings (Saldana, 2021).

Codebook approaches involve the development of a coding framework early in the analytic process, which I then refined as the analysis progressed (Mihas, 2019). Appendix 15 demonstrates the initial code-generating process by using codebook analysis. Then I focus on identifying the key areas (themes) either from the content of the documents or from the codes collected, such as from the section headings or bullet points in the documents (Crabtree and Miller, 1999; King and Brooks, 2018). Appendix

16 provides an example of key area development using codebook analysis. I accomplished this by identifying the essential elements in the documents that serve the study's focus after completing the RTA's familiarisation phase (Braun et al., 2019). Although this approach may limit the depth of analysis, it is considered a pragmatic style of analysis to extract the necessary information without overwhelming the research with irrelevant data (Mihas, 2019; Braun and Clarke, 2020). Roberts et al. (2019) argue that the codebook approach is adaptable because it enables the researcher to determine the level of engagement with the data; thus, the researcher can control the depth of document analysis. This enabled me to generate comprehensive tables highlighting what was mentioned and omitted in the case study documents, allowing me to focus my analysis on the most important aspects to my research and exclude irrelevant document content (King and Brooks, 2018).

Following the procedure outlined above, I conducted three codebook analyses: one for the assessment documents from the three international medical education organisations listed above (3.8.1) and two for each case study assessment document. The codebook also enables researchers to relate their findings to the relevant literature and other data generated for the same study (Roberts et al., 2019; Saldana, 2021). Nadin and Cassell (2004) and Saldana (2021) suggest using matrices to present, compare, and contrast the codebook findings. Thus, I created a matrix to compare each case study's codebook findings to the codebook findings of the three organisations, which I presented in the Findings and Analysis Chapters (Table 7 and Table 15). Moreover, I compiled a third exhaustive matrix that combines the three codebook analyses for the Discussion Chapter (Appendix 17).

Another distinction between the RTA and the codebook is that the codebook does not require quotations from the documents but requires a description of the codes (Mihas, 2019; Roberts et al., 2019). However, I appended sample documents from each case study, as indicated in the documentary analysis in each case study analysis chapter (Appendices 18 to 23).

3.8.2 Interviews and FGs analysis

In the sections that follow, I explain why I used thematic analysis to analyse the interviews and FGs, why I used RTA, and a description of the RTA process.

3.8.2.1 Why thematic analysis?

The interviews and FGs in this research aim to explore how the OSCE is being implemented in SA medical schools. The data collected from the people working on the OSCE provided valuable insights into how the OSCE was designed, conducted, and evaluated. However, a sophisticated analytical approach is required to extract meaning from the available data. Therefore, selecting the most suitable framework for my data

analysis was challenging. I have considered several common qualitative analysis methods to make the most appropriate selection for my philosophical stance and research questions. I examine the applicability of approaches such as content analysis, discourse analysis, ground theory, and thematic analysis to my context, as each would enable me to analyse data differently. Although content analysis is appreciated as a flexible and transparent approach, researchers typically use it to quantitatively analyse visual and printed materials rather than inductively generate meaning (Bryman, 2016). In addition, the content analysis approach is based on post-positivist theoretical assumptions, which are incompatible with the constructivist perspective I embraced (Terry and Hayfield, 2020; Braun and Clarke, 2021).

Another approach is discourse analysis, which looks at linguistic patterns and discursive structure to examine language use, social practises, and power dynamics within data (Cohen et al., 2018). Furthermore, this method focuses on analysing texts that are initially written, such as stories and personal narratives, as opposed to conversations and interviews (Bryman, 2016). Another option is grounded theory, which typically begins with a theoretical framework and employs empirical data to produce a theory with a theoretical generalisation (Silverman, 2017). However, this research has neither a preconceived theoretical framework nor the intent to produce a theory or generalise research findings; rather, it seeks to develop an understanding of the investigated topic and case-to-case transferability (Varpio et al., 2021).

3.8.2.2 Why RTA approach?

Exploration of the thematic analysis approach reveals that it includes several methods, such as framework analysis, coding reliability, codebook analysis, and RTA (Braun and Clarke, 2013; Byrne, 2022). However, considering the research questions and objectives, as well as my theoretical stance, it is apparent that the RTA is the most appropriate approach for several reasons. It provides a clear yet flexible framework that grants a systematic and holistic approach to data, allowing me to generate the meaning and patterns I demand from my data set (Braun and Clarke, 2006; Braun and Clarke, 2021). Braun et al. (2014) and Braun and Clarke (2020) argue that RTA is distinct from the majority of other thematic analysis methods. For example, the themes in RTA are outputs rather than inputs in the analysis process, i.e., they cannot be predicted in advance of data engagement. Additionally, RTA is more aligned with the philosophical perspective I employed in this study, namely constructivism. For instance, the critical meaning discussed by one participant may be more significant than the frequency with which many participants repeat a specific word.

According to Kiger and Varpio (2020) and Byrne (2022), the flexibility of RTA renders it suitable for answering the types of research questions used in my study, i.e., 'how' and 'what' types of questions. Therefore, it is appropriate to comprehend how OSCE is being

implemented in Saudi medical schools and what opportunities and challenges OSCE's adoption in Saudi medical schools offers. Several researchers have lauded RTA as a useful method for analysing data, one that works well under the constructivist philosophical stances I adopt and with a wide variety of qualitative data, such as interviews and FGs (Braun and Clarke, 2006; Guest et al., 2012; Bryman, 2016; Bennett et al., 2019; Kiger and Varpio, 2020; Byrne, 2022). Moreover, comparable studies employ an RTA framework to facilitate data analysis (Campbell et al., 2021; Kua et al., 2022; Shannon et al., 2022).

One important advantage of RTA is that it enables the researcher to be more reflective (Nowell et al., 2017) in that it requires the researcher to make active choices throughout the research. Thus, researchers need to provide ongoing justifications for their choices and reflect on their interpretations of data rather than merely filling the gaps in a procedural manner to get some results without the need for reasoning and reflection (Barrett et al., 2020). This makes the researcher’s positions, values, and experiences prominent, providing a more authoritative dimension to the researcher in analysing the data and actively generating themes.

RTA can also function in inductive and deductive ways to analyse data (Braun and Clarke, 2006). The inductive approach means that I try to start with no previous assumptions and let the data drive the analysis, i.e., by starting from the bottom up and focussing on what participants are saying explicitly or implicitly. By contrast, the deductive approach is driven by a set of questions that I am trying to answer through the data. It is difficult to pick and stick with one of these two approaches because they are deeply intertwined, but it is important to start primarily with one approach while being prepared to use the other when deemed appropriate (Braun and Clarke, 2022; Byrne, 2022). I find the inductive approach to be more plausible as a primary approach for my needs, as I attempt to analyse the data with no preconceived notions, which should allow for deeper and wider coverage of the data. However, being completely inductive would be impossible because I used what the literature suggests to inform my position as well as my experiences in the context and with participants, as discussed further in the ensuring research quality section (3.9) and reflexivity section (8.6).

However, RTA is not without limitations; for example, Flick (2014) criticises the thematic analysis approaches for being influenced by the subjectivity and assumptions of the researchers during data interpretations. Nonetheless, Braun and Clarke (2013) argue that researchers' contextual understanding is a valuable input that has to be considered. However, they must reflect on their own biases and perspectives to promote deeper analysis and generate genuine meanings. Similarly, some researchers claim that researcher bias can diminish the findings of thematic analysis methods (Attride-Stirling, 2001; Guest et al., 2012). According to Varpio et al. (2021), from a positivist perspective, bias is strictly prohibited in research, and the researchers’ personal opinions, choices,

and interpretations are undesirable; however, constructivist researchers use subjectivity to inform their research decisions. Yet, they must consider qualitative researchers' rigour, such as reflexivity, credibility, transparency, coherence, and presentation clarity (Kuper et al., 2008b; Johnson et al., 2020). A common disadvantage of the RTA method is that it is time- and resource-intensive (Byrne, 2022). Braun and Clarke (2022) acknowledge this difficulty and argue that the depth of the analysis, the rich and detailed description, and the iterative relationship between the researcher and the generation and editing of findings are all well worth the time and effort required to generate valuable knowledge.

3.8.2.3 RTA process

This section describes the six-phase framework I used to engage with and analyse the data. This framework was developed by Braun and Clarke (2006) and reviewed by Kiger and Varpio (2020) and Braun and Clarke (2022). However, the process of analysing data using these phases is recursive rather than linear:

1- Familiarisation.

The first step in familiarising myself with the data corpus was to transcribe the interviews and FGs verbatim (Bryman, 2016; Cohen et al., 2018). It allowed me to immerse myself in the data and kept me engaged so I could comprehend the depth and breadth of the data. Appendices 24 and 25 show sample interview and FG transcriptions, respectively. At the beginning of each Analysis Chapter, I provided a table briefly describing each participant, their pseudonym, and the length of the interviews and FGs (Table 5 for case study A and Table 13 for case study B).

The second layer of contact with the data is the careful and active reading and rereading of each data set, which enabled me to take notes and search for initial concepts and meanings. To facilitate future analysis, I kept the data set for each case study in separate folders in NVivo.

2- Coding.

In this phase, I worked to generate codes by rereading the transcripts, this time with a focus on generating meaning 'codes' from each segment in the transcript. I accomplished this by looking at each segment or sentence, attempting to comprehend its meaning, searching for semantic or latent meaning, and then assigning them a brief representative 'code' that can accurately reflect the original meaning. I coded the entire transcript while paying special attention to relevant information. I could assign multiple codes to each segment, which would later be useful for theme generation. I chose to make my codes data-driven or inductive rather than theory-driven, as I believe these capture more original meaning. I included the fieldnote at this stage to help clarify the context and inform the coding processes.

I followed Creswell's (2014) and Bryman's (2016) recommendations for using Computer-assisted qualitative data analysis Software (CAQDAS), so I used NVivo™ with all the

data that I collected. Appendix 26 provides screenshot depicting the code generation process using NVivo. It is beneficial because it facilitates data organisation and analysis, enables me to code text while reading, and facilitates the retrieval of specific information, particularly for large data sets. Appendix 27 includes a list of the initial codes from the four interviews and four FGs of case study A.

3- Initial themes generation.

During this phase, I concentrated on constructing initial themes by searching for all codes and attempting to identify codes with shared patterns and similar characteristics. Then, clustering them into groups that appear to share similar concepts, with each group representing a potential subtheme. I used a thematic map and tables to visualise and classify all codes according to their appropriate themes. After that, I worked to generate themes from similar subthemes that incorporate numerous codes with interconnected meanings. Appendix 28 demonstrates one of the earliest cycles for developing themes and subthemes. Additionally, during phases two and three, I extracted interesting statements and potential quotations from the transcripts and compiled them in one location for use as evidence to support my themes and findings during the writing phase. I used Microsoft OneNote™ to facilitate this step and make selecting the quotation more practical. Appendix 29 shows a sample of my strategy for collecting the best quotations during the RTA analysis.

4- Themes developing and reviewing.

This phase is crucial for developing themes and ensuring that the themes generated are consistent with the identified codes and the entire dataset. This necessitates rereading the selected extracts and comparing them to the original data corpus to ensure that my themes are accurate and reflective, i.e., checking their validity to ensure that they accurately represent the original data. I also re-examined the generated themes, removed unfit subthemes and themes, and added new ones that became noticeable after the reviews. This phase involves reviewing them and checking that the generated themes have both internal and external homogeneity (Patton, 2015). Internal homogeneity refers to the notion that the content of the themes must be clearly linked to a central organising concept, whereas external heterogeneity denotes that there is no significant overlap between the themes that blurs the lines that distinguish them from one another.

Another strategy I used was to revisit the themes and subthemes regularly (e.g., every few weeks) to ensure that they were representative of the data and that they highlighted the key findings (Flick, 2014). Performing this *'time spacing'* revealed that most subthemes and themes were accurate; however, it also revealed that a few of them appeared to require further development or even modifications to better fit the theme content. As a result, I made some significant revisions. For example, I used to have three themes with seventeen subthemes, but with further development and revision, I made

them four themes with ten subthemes, which appear to have a more coherent and convincing structure. Appendix 30 displays one of the later cycles of theme and subtheme development.

5- Themes refining and defining.

This stage entails refining, defining, and naming themes and subthemes by providing a clear description or brief synopsis of what I mean by each theme and subtheme to ensure that the names I use accurately reflect their content. I aimed to ensure the face validity of the themes and subthemes, i.e., upon reading the title, readers should have a good idea of what the topic is about. I also focused on delineating the themes and subthemes and avoiding unnecessary overlap between them. It is vital to convey the scope of each theme in its description, so when the scope is excessively broad, I consider splitting them into two. In some cases, I deemed it more appropriate to subdivide the main theme into subthemes or the subtheme into subheadings. However, when the scope is too narrow, I might consider making the theme a sub-theme under another theme. Figure 13 shows the basic hierarchical structure of the themes.

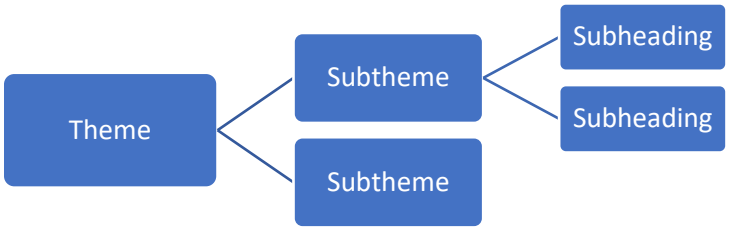


Figure 13. The basic hierarchical structure of the themes.

6- Writing-up.

Although writing is an integral process throughout all RTA phases, including notes and reflexive journaling, there comes a time when most of the work focuses on formal writing to produce the final analysis. During the writing phase, I attempted to present my analysis coherently and persuasively. This involves selecting compelling and pronounced extracts to serve as references and evidence to support the validity of my analysis. I combined the quotes with interpretation and discussion to further explain the themes' meaning. In addition, I provided an analytical argument demonstrating the relationship between the data and my findings, on the one hand, and my findings in relation to the literature and research questions on the other. Figure 14 summarises the RTA analytical process.

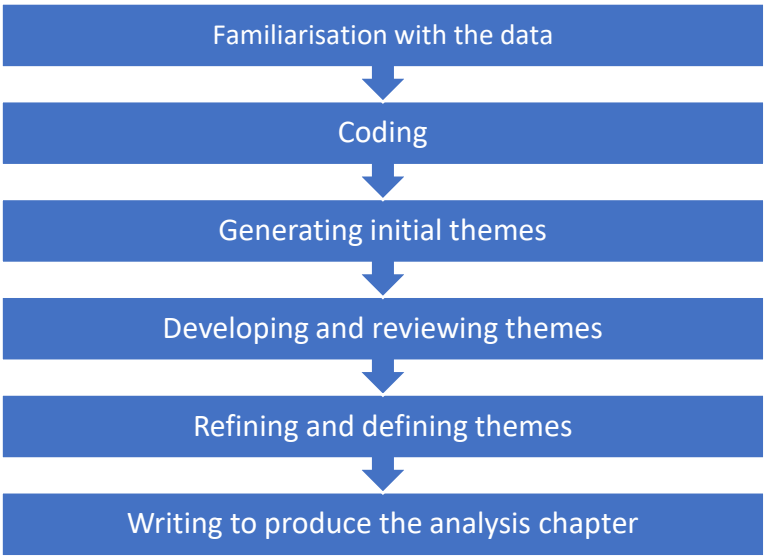


Figure 14. A summary of the RTA analytical process.

3.8.3 Notes on my approaches to analysis

The RTA approach of Braun and Clarke (2022) motivates me because it encourages me to reflect while analysing and writing. Reflexivity means I, as a researcher, should continue to question and reflect on my choices throughout the analysis process (Braun and Clarke, 2020). For example, I kept asking myself, is there a different way to look at this? It is akin to an internal debate. The notion of reflexivity also encourages me to bring my thoughts and ideas to research. I have some experience with the assessment system and OSCE practice in SA, which I used to make the results and discussion more meaningful. Nonetheless, this could be a drawback, as I may be so familiar with the OSCE and assessment practice in SA that I overlook significant detail. However, this is where my supervisors ask me to clarify and reflect on my writing to make the results and discussion more credible.

RTA and codebook procedures were iterative, seeking patterns by adding, deleting, and modifying codes and themes until they aligned with the research focus and could answer the research questions. As Braun et al. (2014) describe, I found RTA analysis helpful in generating deep meanings from data that appeared disparate at first glance. Furthermore, this process aided in generating findings that inform OSCE practice in SA schools. The RTA and codebook allowed me to use the criteria of good assessment (discussed in Chapters 1 and 2) and the literature review as a theoretical framework to inform the findings and facilitate their link to the literature.

Due to the adaptability of the RTA and codebook, I was able to analyse each case study separately; however, because both are designed to investigate the same issue, the themes and most subthemes are identical. Nevertheless, each case's reality and context determine each theme's content and description. I subsequently utilised Stake (2006) and Yin's (2018) recommendations to combine the two case studies in the Discussion Chapter by employing a cross-case synthesis model. This entails comparing, contrasting, and connecting the findings of various instruments across case studies to answer the research question and address its objectives (Yazan, 2015; Cohen et al., 2018).

3.9 Ensuring research quality

The quality and rigour of any research are of the utmost importance for determining the trustworthiness and credibility of the research process and findings (Lincoln and Guba, 1985). Many authors, including Elliott et al. (1999), Kuper et al. (2008b), Tavakol and Sandars (2014), and Cleland (2022), emphasise the distinctions between qualitative and quantitative research quality criteria. However, there are differing views on what constitutes a quality standard in qualitative research (Elliott et al., 1999; Yardley, 2000; Cohen and Crabtree, 2008; Kuper et al., 2008a; Miyata and Kai, 2009; Stige et al., 2009; Tracy, 2010; Arthur et al., 2012; Miles et al., 2018; Johnson et al., 2020). Although different researchers used different quality criteria, on reflection, I find them to be considerably interconnected and overlapping, with a few subtle differences. Therefore, I considered criteria accepted by most of the authors and appropriate for the constructivist paradigm to ensure the high-quality of this research (Finlay, 2006). Those criteria involve reflexivity, sensitivity to context, transparency, coherence, clarity, transferability, triangulation, credibility, impact, importance, and resonance with readers.

Reflexivity is important in qualitative research because the researcher's background, assumptions, and experience may influence his or her decisions when dealing with data and producing research findings (Yardley, 2000; Kitto et al., 2008; Olmos-Vega et al., 2022). Flick (2014) argues that researchers cannot isolate themselves from the research; therefore, they must be reflexive, transparent, and aware of their influence on data collection and analysis. I acknowledged my background as a medical doctor who works as a medical educator and researcher in a medical school in SA that has a similar context to the case studies in this research. In addition, I have been an OSCE assessor, so these experiences have provided me with knowledge of the medical education system and environment in SA. My experience can influence this project, as I have collected the data in a familiar environment. I am an insider researcher with extensive knowledge of the research context before beginning. The advantage is that I understand the culture because I am a part of it, and I may have better access to medical schools for data collection. However, I am cognisant of my situation, which I believe prevented me from

overlooking alternative options because I am familiar with the surrounding environment. I was aware that, based on my prior knowledge, I could make assumptions about the data or individuals without probing them.

It is hoped that being reflexive and openly disclosing my situation kept unavoidable bias in check, utilising the discussion with my supervisors and peers to develop balanced opinions and decisions as well as preventing me from omitting important details (Braun and Clarke, 2022). Another approach I used was reflexive journaling, which I used as a self-critical account throughout the research process (Sudirman et al., 2021; Braun and Clarke, 2022). The reflective journal assisted me in recording my thoughts and understanding for each step of the research to keep me aware of my positionality, i.e., as a medical educator who is familiar with the SA context and remind me of the circumstances and justification for every decision I made. These measures assisted me in ensuring that I had not overlooked any detail that could have been missed due to my familiarity with the context or even caused an unwanted impact on the results. In the Conclusions Chapter (Section 8.6), I provided a detailed reflexive account of the entire thesis, following Olmos-Vega et al. (2022) recommendation of providing personal, interpersonal, methodological, and contextual reflexivity.

Yardley (2000) and Finlay (2006) argue that sensitivity to context or caring is a quality criterion that qualitative researchers should consider, which includes contextual, theoretical, and sociocultural sensitivity. In the Introduction Chapter, I addressed contextual sensitivity by providing a description of the Saudi medical education context in which this study was conducted (Stige et al., 2009). In the Literature Review Chapter, I addressed the theoretical sensitivity by exploring similar studies on OSCE quality, OSCE operations in new locations, and OSCE operations in SA (Johnson et al., 2020). These studies inform this research by assisting in the design of the research methodology and methods, the development of a more in-depth analysis, and the discussion of my research findings in relation to the broader literature. I have also demonstrated sociocultural sensitivity with research participants by providing a reflective interpersonal account in the Conclusions (8.6.2), providing them with all the information they may need as participants, and obtaining ethical approval to conduct the research (Finlay, 2006; Tracy, 2010).

Transparency is an important quality criterion that refers to the openness and explicitness of reporting the research process in a clear manner that allows others to evaluate different research aspects (Yardley, 2000; Finlay, 2006; Arthur et al., 2012; Johnson et al., 2020). I attempted to accomplish this by providing a thick description and background information so that readers could judge for themselves the transferability, credibility, and reliability of the data collected and research results (Yardley, 2000; Kuper et al., 2008a; Bryman, 2016). I endeavoured to adhere to this principle throughout the research; for example, I have articulated my ontological and epistemological stances,

detailed the data collection process, including sampling and recruitment, and justified the selection of data analysis approaches and processes. Elliott et al. (1999) emphasise the importance of *'grounding in example',* which I accounted for by including appendices that illustrate the hidden work of the PhD. For example, I tabulated the literature review references (Appendices 5, 6, and 7), provided examples from the interviews and FGs transcripts as well as from the assessment documents (Appendices 18 to 25), and presented examples that demonstrate how I worked with the data to generate the codes and themes (Appendices 15, 16, and 26 to 30). To increase transparency, I present a detailed account of the research's limitations and strengths in the Conclusions Chapter (8.5) (Stige et al., 2009).

Some authors emphasise coherence and clarity to maintain logical consistency across various parts of the research (Yardley, 2000; Finlay, 2006; Cohen and Crabtree, 2008). I made an effort to maintain consistency and establish logical and clear connections between the research questions, objectives, research philosophy, sampling, data collection methods, data analysis methods, data interpretation, and recommendations (Kuper et al., 2008a; Johnson et al., 2020). For instance, I used evidence from the data and pertinent literature to substantiate my interpretations and arguments. In addition, I refrained from using quantitative data collection methods such as questionnaires and positivist principles such as generalizability; instead, I used transferability (Varpio et al., 2021). Transferability is the criteria that I considered to enable readers to determine the possibility of transferring and applying my findings in other contexts (Kitto et al., 2008; Miyata and Kai, 2009; Tracy, 2010; Miles et al., 2018). Transferability is largely supported by transparency, as providing a clear and adequate description of each research step enables other researchers to evaluate the applicability of the research's findings to comparable settings or emulate the research process (Johnson et al., 2020).

I also employed triangulation to increase the credibility of my research by collecting data from multiple sources (i.e., academic leaders, assessment specialists, OSCE designers, OSCE writers, and OSCE examiners) using multiple methods (i.e., interviews, FGs, and documents) (Elliott et al., 1999; Cohen and Crabtree, 2008; Kitto et al., 2008). To enhance the credibility and accuracy of the data, I asked the participants for clarification when I did not understand their points, and I rephrased my questions when I believed they had misunderstood me (King et al., 2018). I also made certain that no additional information that the participants could provide to inform the research by asking at the end of each interview or FG if they had anything to add or elaborate on (Cohen et al., 2018). In an effort to further establish the credibility of the research, I attempted to provide plausible and justifiable explanations for all the decisions I made throughout the research (Finlay, 2006). I also spent significant time conducting the analysis by reflecting on the data to develop fresh perspectives to ensure that I had achieved credible and significant findings (Braun and Clarke, 2022).

The impact and importance of the research are another way to judge research quality (Yardley, 2000; Finlay, 2006; Stige et al., 2009; Miles et al., 2018). Throughout this research, I argued the importance of quality OSCE assessment for all stakeholders; therefore, this research aimed to understand how it is implemented within particular contexts and develop recommendations to capitalise on opportunities and overcome challenges for quality OSCE, as detailed in the Conclusions Chapter (8.4).

Elliott et al. (1999), Finlay (2006), and Tracy (2010) agree that the research findings and analysis should resonate with readers as a quality measure. This means that it should accurately reflect the subject matter, represent the research experience, and present the research clearly to the intended audience. Therefore, I utilised the feedback I received from supervisors, peers, and others at international conferences and local events where I presented my work (i.e., oral presentations and posters) to improve the research presentation and make my academic language and claims more relevant to the target audience (Braun and Clarke, 2022). Appendices 31, 32, and 33 offer examples of the presentations and publications that arose from this research and contributed to its development.

3.10 Summary

In this chapter, I presented the constructivist philosophical stance underpinning this research design. Besides that, I outlined the methodological structure of designing the case studies and clarified why I chose these particular data collection techniques, namely interviews, FGs, and documents. In addition, I addressed the process of sampling and recruitment, ethical considerations, and pilot testing. Then, I described the data collection procedure for the two medical school case studies. I also explained why I used RTA to analyse interviews and FGs and codebook analysis to analyse documents. Lastly, I described the quality criteria that I adhered to throughout the research.

Chapter 4 Findings and Analysis—Case Study A

4.1 Introduction

This research aims to answer two questions: how is the OSCE being implemented in Saudi medical schools, and what are the opportunities and challenges offered by adopting it? Therefore, I collected data from two case studies (A and B) from two medical schools in Saudi Arabia (SA) to answer them. Because each school has a unique context, as discussed in the Methodology and Methods Chapter, the results of each case study will be presented in their own chapter. In this chapter, I will present the findings of the first case study—case study A, and in the next chapter, I will present the findings of the second case study—case study B.

Case study A is an example of, and shares the context with, many medical schools in SA, as discussed in the Methodology Chapter (3.3.1.1). The main characteristics of this case study are that it is a newly established, publicly funded medical school with no accreditation or partnerships with other medical schools. The findings are derived from the analysis of fifteen assessment documents, four interviews, and four FGs. The four interviews were with an academic leader and two assessment specialists, and the four FGs were with OSCE designers, writers, and examiners; Table 5 details the data sets.

I will present the document analysis results using the codebook analysis approach (Mihas, 2019; Roberts et al., 2019); second, I will analyse the interviews and FGs. The analysis of the interviews and FGs allows me to structure the findings into four overarching themes, each with some subthemes. Using the Reflexive Thematic Analysis (RTA) approach (Braun and Clarke, 2006; Braun and Clarke, 2022), themes and subthemes were generated after deep engagement with the data sets. More information on the analytical approaches used can be found in Methodology Chapter (Section 3.8).

Table 5. Details of the documents, interviews and FGs conducted for case study A.

#	Data set and participants pseudonym		Participant's description	Length
1	**OSCE documents**	15	See Table 6 for more details	174 pages
2	**Interview 1** (Interview1–AL)	1	Academic leader (AL) 1 Four years in academic service	37 minutes
3	**Interview 2** (Interview2–AL)	1	Academic leader (AL) 2 Two years in academic service	46 minutes
4	**Interview 3** (Interview3–AS)	1	Assessment specialist (AS) 1 Seven years in academic service	38 minutes
5	**Interview 4** (Interview4–AS)	1	Assessment specialist (AS) 2 Twenty years in academic service	39 minutes
6	**FG 1–Designers** (FG1–D)	5	Five female OSCE designers and station writers (From 4 different courses: 1. General Surgery, 2. Psychiatry, 3. Obstetrics and Gynaecology, 4 and 5. Paediatric)	103 minutes
7	**FG 2–Designers** (FG2–D)	5	Five male OSCE designers and station writers (From 5 different courses: 1. General Surgery, 2. Paediatric, 3. Orthopaedics, 4. Otorhinolaryngology, and 5. Ophthalmology)	79 minutes
8	**FG 3–Examiners** (FG3–E)	6	Six female OSCE examiners (From 5 different courses: 1. Internal Medicine - Neurology, 2. Family Medicine, 3. Obstetrics and Gynaecology, 4. General Surgery, 5 and 6. Paediatric)	78 minutes
9	**FG 4–Examiners** (FG4–E)	6	Six male OSCE examiners (From 5 different courses: 1. Internal Medicine, 2. Internal Medicine - Dermatology, 3. Psychiatry, 4. Paediatric, 5 and 6. General Surgery)	89 minutes
Total 9 data sets		**26 participants and 15 documents**		**509 minutes**

4.2 Documentary findings and analysis

This section is a documentary analysis of the available documents from this medical school, as discussed in Methodology Chapter (3.8.1).

4.2.1 Findings of the documents

For the document analysis, I used the codebook analysis approach to analyse the documentary evidence from School A (Mihas, 2019; Roberts et al., 2019; Braun and Clarke, 2020). I collected fifteen documents that I classified into three categories: 1) assessment policy—two documents; 2) examples of OSCE marking schemes—nine documents; and 3) other assessment documents that proved not relevant to my study—four documents. As displayed in Table 6, I conducted a comprehensive review of the documents and summarised the relevant content in each document, along with a commentary on the contents of each. I was unable to append all documents because they are lengthy, and doing so would reveal the identity of the medical school. Therefore, I chose to provide examples of documents that will not reveal their identity (Appendices 18, 19, and 20).

Then I performed the codebook analysis, as shown in Table 7. I compared the relevant content of these documents with the assessment documents and guidance from three international organisations concerning assessment implementation in medical schools. The three organisations that I used their criteria and frameworks are the Ottawa Conference, AMEE ASPIRE-to-excellence award in assessment, and WFME, more details are in Methodology Chapter (3.8.1) (Boursicot et al., 2011; WFME, 2015; Boursicot et al., 2020; WFME, 2020; AMEE, 2023).

Table 6. This table provides an overview of the case study A assessment documents and each document's relevant content with a description.

Doc. No.	Document Title	Pages	Relevant Content	Description
1	Assessment and Evaluation Guide (Dated 2013) (Appendix 18 contains some excerpts from this document)	105	It is a guide for the assessment system in this school; the relevant content includes content about: • Criteria of good assessment. • Role of the assessment unit. • Brief OSCE description. • Designing examination blueprint. • Designing marking schemes.	It is a practical guide and reference for this school's assessment system and assessment tools. It is lengthy and detailed in some places but concise and negligent in others. The OSCE content is very limited, mentioning the OSCE on one page, providing a short description, advantages, disadvantages, and when OSCE can be used. Hence this cannot be regarded as an in-depth guide to the OSCE. It lacks key information such as the purpose of the OSCE; the need for exam organisers, station writers, assessors, and SPs training; and how to construct the stations. Notes: •Good things to identify: *Importance of well-designed assessment strategies p2, p39. *Following good assessment criteria p2/9/10. *Blueprinting p23/24. *Encouraging psychometric analysis p43/81/82/83/84, *encourage using rubric p43/44/47/52, and criterion-referenced method for all assessments p43. *Encouraging feedback to students p47/58/59/60/70. • Inconsistent with the literature: *Only 3-5 stations in the final OSCE exam p39 but on p30 said 15-20 stations. *Only 3-5 mins for each station p39 but in p30 said 15 min. *Underestimate the difficulty of designing valid scoring checklists p33. •Limitations identified: *No proper venue p40/41. *Lake of staff (examiners) p40. *The need for different types of models (manikins) p40.

				•Explicit regulations: *Follow university regulations in the grading system (pass score is fixed at 60 out of 100) *The roles of the Examination Committee in some detail p68/69/70.
2	Assessment Policy (Dated 2017) (Appendix 19 contains some excerpts from this document)	45	It is a guide for the assessment policy in this school; the relevant content includes content about: • Practical-based assessment (Chapter 3) • Quality assurance (Chapter 5)	Provides general policies governing assessment activities at this school. It briefly mentions eight types of performance assessment, namely OSCE, OSPE, 360-degree evaluation, SDOT, Mini-CEX, COT, LEP, and P-MEX. Notes: • Inconsistent with the literature description of OSCE: *Definition of OSPE as this overlaps with OSCE p20. *Each station has two independent examiners p21. * Inconsistent OSCE naming "Objective Specific Clinical Examination". •Suggestions with no evidence-based but not used (their practice is different than what has been written): *Video recording for all stations p21. *Rest station after each station p20. *Assessors not allowed to be in the station p21. *Two independent examiners in each station. •Good practice mentioned but not used: *Use of scoring rubric p22. *No interaction between examiner and examinee p21. *Exam preparation by course organiser p31. *Post-hoc analysis but nothing specific for the OSCE. *Encourage question bank. • Overlooked issues: *No min/max total exam time. *Station duration. *Number of stations. *Importance of conducting formative OSCEs. *Training of assessors and SPs. There is an inconsistency between terms used in this document, e.g., Angoff method & minimum performance level.
3 to 10	Six surgery checklists and Two orthopaedics checklists	10 in total	Provide an idea about their marking scheme	It is a simple binary checklist with a mark for each task, and the whole station mark is 20. There is no rubric description and no further details to help the assessors' decision-making. However, some tasks provide some details. Some tasks have a different weight than other tasks.

	Two examples in Appendix 20			This is different from the suggestion of the assessment unit policy of using a detailed rubric scoring scheme.
11	One marking rubric form	1	Generic, so it is not very relevant.	No narrative description of how to assess performance. It only shows a table of 5 Likert scales to describe performance for any task by Excellent, very good, good, fair, and poor.
12 & 13	Online Assessment (2 documents)	5 in total	Irrelevant - not relevant to my focus	It indicates the importance of using a wide range of online assessment tools and mentions some online assessment tools examples.
14 & 15	Exam Policy (for students & staff) (2 documents)	8 in total	Irrelevant - not relevant to my focus	It provides general managerial rules, such as exam attendance, when they can leave the exam hall and official non-academic rules.

1.2.1.1 The codebook analysis matrix

Using codebook analysis, I compiled a set of codes derived from the three international organisations' assessment documents and guides. Then, I created a comprehensive matrix comparing all codes found in the resources of the three guides to those found in this medical school's documents. Following this approach, four key areas emerged: 1) assessment policy, 2) criteria for good assessment, 3) use a system of assessment, and 4) educational resources for assessment—each theme contains multiple codes, fifty-eight in all. Table 7 presents the codebook analysis matrix. Most of the codes are present in all documents; however, as shown in Table 7, a few are missing from this school in comparison to the literature codes. Appendix 15 contains a worked example from the documentary material that depicts the code generation process. In addition, the table in Appendix 16 demonstrates how the codes were refined and then combined into a single area.

Table 7. This matrix illustrates the codebook analysis results for case study A's assessment documents in relation to the three organisations' assessment documents.
The underlined points indicate a discrepancy between this school's documents and the literature.

Key areas	**Codes from literature** Ottawa conference consensus + ASPIRE criteria + WFME standards	**Codes from this medical school** The assessment documents of the case study A	**General guidance for the codes**
Assessment policy	• Assessment policy should be responsive to the context	• Assessment Guide 'detailed account for assessment policy' • Founded assessment unit and examination committee	The school assessment policy should govern and tailor the assessment according to its context and available resources.
	• Specified educational outcomes	A lack of evidence in the documents sampled.	The assessment plan should indicate specific educational outcomes expected from the assessment process.
	• Designing & conducting the OSCE ○ Marking scheme ○ Writing the stations	• Designing & conducting the OSCE ○ Marking scheme ○ Writing the instructions ○ Stations with no assessors (only written questions) ○ Dependent stations ○ No assessor–assessee interaction	The elements and steps to design, implement, and review the OSCE.
	• Appeal policy	• Appeal policy	Detailed account for petition and challenge for the assessment results.
	• Remediation process	• Remediation policy	Detailed account for the re-sitting of the exams for students.
Criteria of good assessment	• Validity ○ Content blueprinting	• Validity ○ Content blueprinting	The assessment measures what it is supposed to measure and reflects the curriculum.
	• Reliability ○ Number of the stations ○ Length of the stations ○ One assessor is recommended	• Reliability ○ Number of the stations ○ Length of the stations ○ Two assessors are recommended	When the same assessment is introduced again in similar circumstances, it should produce similar results. Multiple issues mentioned in the documents contribute to the reliability of the OSCE, such as sampling for the

	○ Standard-setting (criterion-referenced) ○ Correlation with other exams ○ Psychometric analyses	○ Standard-setting (fixed-referenced) ○ Parallel stations ○ Statistical analysis	stations, ensuring adequate testing time, considering standard-setting, and managing parallel stations.
	• Fairness ○ Test security ○ Examiner training	• Fairness ○ Test security ○ Controlling application ○ Oversight the exam	The OSCE should use a fair grading system and be free of deliberate bias, cheating, or any other issue that gives advantages to some students over others.
	• Educational effect ○ Feedback to the students ○ Formative exams	• Educational effect ○ Feedback to the students ○ Formative exams	The educational benefit that students gain from taking the OSCE. For example, providing feedback for the students and providing formative OSCEs.
	• Catalytic effect ○ Feedback to the school and teaching system ○ Quality assurance	• Catalytic effect ○ Feedback to the school and teaching system ○ Quality assurance	The motivation that all stakeholders gain from conducting the OSCE. For example, the feedback that the school received to support and enhance future education, reviewing the implementation process for quality assurance purposes, and improving future implementation.
	• Feasibility	• Feasibility ○ Station bank	The assessment should be practical in terms of time, funds, and workload in a given context.
	• Acceptability	• Acceptability	Different stockholders—faculty and students—recognise assessment importance and credibility regarding its content and process.
	• Equivalence	A lack of evidence in the documents sampled.	The assessment results lead to similar or equivalent decisions across equivalent cohorts or institutions.
	Not explicitly stated	• Authenticity	The OSCE should reflect and mimic real-life situations.
Use a system of assessment	• Comprehensive ○ Use multiple tools for assessments	• Comprehensive ○ Use multiple tools for assessments	The OSCE should not be used alone as a clinical assessment tool, so various assessment methods need to be employed.
	• Coherent	A lack of evidence in the documents sampled.	It emphasises that different assessment tools in the assessment system should be consistent and aligned around the same purpose.
	• Continuous	A lack of evidence in the documents sampled.	The assessment system needs to be continuously improved and evaluated against quality criteria.
	• Purpose-driven ○ Formative and summative	• Purpose-driven ○ Formative and summative	It indicates the assessment system should include formative and summative exams to improve students

	○ Assessment role		to graduate competent doctors able to contribute to society's health and mind patient safety.
	• Transparent ○ Utility indicators ○ Assessment supported by the assessment literature	• Transparent	It required the school to make its assessment procedures accessible to all stakeholders to endorse accountability and integrity principles. Transparency should encourage the school to use evidence from the literature to make good assessment decisions.
	• Acceptable and appropriate	A lack of evidence in the documents sampled.	The assessment system should be designed to be acceptable and appropriate for all stakeholders.
	• Feasible	A lack of evidence in the documents sampled.	The assessment system needs to be efficient and practical in terms of the fund, time, and workload regarding the school context.
	• Quality control ○ Feedback to school ○ Evaluation and analysis of performance ○ Continuous improvement ○ Scholarship promotion	• Quality assurance ○ Feedback to school ○ Reviewing and evaluation	The school needs to ensure its assessment quality by evaluating current applications to improve future applications. This can be done by evaluating the assessment utility, which is a combination of the items mentioned in the criteria of good assessment.
Educational resources for assessment	• Human resources ○ Experts for design and support ○ Examiners ○ Patients and SPs	• Human resources ○ Mention the lack of staff	The assessment documents emphasise the need for assessment expertise, examiners, patients, and SPs. These school documents mentioned a lack of them.
	• Stakeholder needs ○ Faculty development ○ Examinees needs ○ Healthcare system and regulators needs ○ Patients needs	• Provide support for the staff ○ Faculty development	All stakeholders should receive enough familiarisation and training for the OSCE as a tool for assessment.
	• Physical resources ○ Models ○ Venue • Fund	• Resources ○ Mention the lack of models ○ Mention the limitations of the space	The need for fulfilling assessment physical and financial requirements. These school documents mentioned a lack of them.

4.2.2 Analysis of the documents' findings

This section compiles and analyses the key findings of the document analyses presented in Table 6 and Table 7. It is structured in accordance with the broad subheadings of the validity framework workbook for the OSCE described by Boursicot et al. (2019, 2022).

4.2.2.1 General comments on the assessment documents

Table 7 highlights that this school's documents establish almost all key issues that need to be documented in their assessment documents, with a few exceptions. The themes and codes in this school's documents are similar to those in the three organisations' documents. However, some points are not mentioned in the documents, such as the coherence, continuity, acceptability, and feasibility of the assessment system. To find out whether the organisation is aware of these issues or not, it may be necessary to look at data from other sources.

Even though the documents contain some information about the OSCE, they do not provide a comprehensive description of its conduct and implementation. Moreover, the Assessment and Evaluation Guide identified some logistical limitations to the school's OSCE implementation. It mentioned the absence of a proper venue for OSCE exams, the lack of staff and examiners, and the need for different types of manikins for different OSCE stations. The author indicates that these limitations are due to the school being newly established and thus requiring additional resources to meet the needs of this demanding assessment.

4.2.2.2 Intended use of the OSCE in the documents

The intended purpose of the OSCE is not stated in any documents made available to me. However, it indicates the competencies that this tool can assess, which include history taking, physical examination, simple procedures, patient management problems, communication, and attitude. Nevertheless, there is no mention of how the OSCE is integrated into the wider assessment system. However, the documents highlight that the performance assessment programme should employ various instruments such as OSCE, Mini-CEX, DOPS, and short and long-case examinations. Also, it indicates the need for both summative and formative exams, with no mention of when and how each type of OSCE should be used.

4.2.2.3 Relevant data to support the intended use of the OSCE

Planning of content

The documents emphasise the importance of designing a valid, reliable, and successful OSCE. While there are limited instructions for how to design the OSCE, the documents do not provide complete descriptions of how the OSCE should be structured following

evidence-based practises. The document stresses the significance of blueprinting to ensure appropriate sampling across all clinical domains and skills. Then, all exam blueprints should be approved by the examination committee to ensure their quality before implementation.

Internal structure

The documents specify some criteria to consider during the design stage, but there are some inconsistencies. For example, the Assessment and Evaluation Guide recommends only three to five stations for the final OSCE exam, but elsewhere, it recommends fifteen to twenty. Similarly, the guide suggests that three to five minutes is enough time at each station, but it recommends fifteen minutes elsewhere. Furthermore, there is no process for piloting stations to ensure task-time agreement. The guide details and encourages the design of appropriate marking rubrics for each station; however, the marking schemes they provided me as examples do not reflect the guide's recommendations (Appendix 20).

Response process

To conduct a fair OSCE, the documents encourage the use of two independent examiners at each station to provide accurate assessments for students. However, the literature does not support this approach, as the two independent examiners would not add any significant advantage compared to one examiner (van der Vleuten, 1996; Harden et al., 2015). The documents also emphasise the role of training and provide accurate information sheets for different stakeholders to implement a successful OSCE. However, I was unable to locate specific information on the training processes, length, instructions, or if OSCE designers, examiners, and SPs were required to attend.

Consequences of the OSCE

The standard setting recommended in the Assessment Policy is the criterion-referenced approach, such as the Angoff method. However, the Assessment and Evaluation Guide states that the cut score is sixty (fixed-referenced) in all assessments and courses, which is actually practised in this school. Homer et al. (2016) and Park et al. (2021) argued that the BRM is the recommended method to set the cut score, highlighting the inconsistency between this school practice and the evidence-based practice. In addition, there is a lack of details in calculating pass grades; for instance, it is unclear whether students must pass each station or a minimum number of stations to pass the entire OSCE examination. Likewise, the documents do not specify the regulations and provisions for those who fail the OSCE.

Psychometric analysis is encouraged in the documents to promote quality control and evaluate the implementation by suggesting the use of metrics such as the difficulty index, the discrimination index, and the reliability coefficient. Furthermore, the documents

emphasised the significance of providing students with feedback and outlined the criteria for quality feedback in order to improve the educational value of the OSCE (Boud, 2015; Hattie and Clarke, 2018). However, the documents do not describe the procedure for providing written or verbal feedback during or after the exam.

4.2.3 Summary

Evidence demonstrates that this institution attempts to develop OSCEs that are reliable, fair, and educationally valuable. While the assessment documents have corresponded well with the recommendations and criteria of the three organisations in many areas, this school's documents lack internal consistency in places and mismatch with the literature guidance in others. For example, the use of a fixed-referenced cut score and the recommendation to employ two independent examiners at each station.

4.3 Interviews and focus groups findings and analysis

Using RTA to analyse this case study's interviews and FGs allowed me to generate four main themes (which are shared by both case studies, as explained in Section 3.8.3):

1) Institutional and assessment culture: The influence of institutional and assessment culture on the OSCE implementation.
2) Faculty expertise and practices: The effect of faculty expertise and practice on the current OSCE implementation.
3) OSCE quality and design: The current OSCE procedures impact on the OSCE quality and design.
4) Resources and infrastructure setup: The impact of resources availability on the OSCE implementation.

Table 8 summarises the themes and subthemes derived from case study A's interviews and FGs. Each theme consists of multiple subthemes, a total of ten, some of which have subheadings, summarised in Appendix 34. Due to the integrative nature of the qualitative case study analysis, the themes and subthemes may have limited overlap. Therefore, each theme or subtheme must be considered in relation to the others, as their analysis in isolation is untenable. Therefore, at the end of this chapter, an overarching theme will connect the concepts discovered and analysed in this case study.

Table 8. A summary of the themes and subthemes derived from the interviews and FGs.

#	Themes	Subthemes
1	Institutional and assessment culture	The nature of communication and collaboration
		The Influence of assessment regulations
		The desire for improvement
2	Faculty expertise and practices	Faculty awareness of the OSCE
		Faculty efforts in planning and preparation
3	OSCE quality and design	Attention to quality
		Theory-practice gap
		Reliability issues
4	Resources and infrastructure setup	Human resources
		Educational and logistical resources

4.3.1 First theme—Institutional and assessment culture

This is the first theme in which I aim to draw a general picture of the institutional and assessment culture that affects the OSCE in this particular medical school. According to Holzweiss et al. (2016) and Simper et al. (2022), the institutional culture has a significant impact on the assessment culture in educational institutions. This theme explores the factors and issues shaping the culture of this institution's assessment and influencing the OSCE implementation. I divided the aspects that affect how the OSCE is being utilised in this school into three subthemes, Table 9.

Table 9. The subthemes of the first theme and their description.

Subthemes	Description
The nature of communication and collaboration	Explores the aspects of internal and external communication and collaboration among various stakeholders that affect OSCE implementation
The Influence of assessment regulations	Examines the influence of external and internal regulations on the OSCE's implementation
The desire for improvement	Demonstrates the impact of being a newly established medical school and the desire to improve their OSCE in the future

4.3.1.1 The nature of communication and collaboration

This subtheme discusses the nature of this school's communication and collaboration with different stakeholders. It investigates this issue within the school itself, the neighbour medical schools, and the neighbour hospitals. This subtheme concerns internal and external communication and collaboration indicators, which may affect this school's OSCE implementation.

Internal communication

This section shows the communication issues between staff and faculty to staff. This subtheme illustrates the nature of the relationships in this medical school and their impact on the OSCE.

Participants highlighted a number of issues that demonstrate a lack of communication among faculty. For example, one participant reflected on her feelings regarding asking questions and communicating ideas.

> *'We always feel asking others—when we face a problem—is difficult, and we feel embarrassed about what other people will think about us when we do not know. We do not want to appear like we do not have the experience. But asking colleagues and sharing information could open our minds to new horizons.' (FG1-D5).*

Another OSCE designer demonstrated the lack of experience exchange between the faculty.

> *'I notice from this discussion with colleagues from different departments that I learned a lot from them… So, I wish every course organiser could share his experience with the OSCE and other exams. We can, for example, organise a weekly meeting with one member of staff from each department, and then we can discuss the obstacles and how others have overcome them, and so on.' (FG1-D3).*

One interviewee explained that the unsupportive communication culture prevented her from conducting statistical analysis.

> *'To be honest, I want to do some analysis of students' grades and performance, but other parties, like staff and departments, are not willing to communicate or help experiment with something new. I do not know how to do this, and I receive no support to do so; I think these are limitations.' (FG1-D2).*

This suggests a lack of an effective communication culture at this medical institution, which could hinder the implementation of a better OSCE. According to the quotations above, some staff lack the motivation to attempt new things and therefore prefer to maintain the status quo. In addition, some quotations imply a disregard for the proper practise of adhering to administrative guidance, as in the following quotation.

> *'The assessment unit in the college recommends an exemplar scoring rubric, but it seems difficult to use. So, I designed our own checklist for our course… Frankly speaking, I think the assessment unit members will not like our checklist. I am a member of the assessment unit, and I know that. I believe our checklist in our course is more appropriate than the one they suggest we use.' (FG1-D5).*

The above OSCE designer seems not to prefer discussing this issue with the assessment unit. Likewise, according to the data, many faculty members do not attend faculty development workshops. One participant offered what he believed to be a valid justification:

> *'It was not mandatory, so many staff easily neglect this if they think they have something more important to do like a clinic to attend, a lecture to give, or something else. But they should design their schedules the other way around.' (FG2-D5).*

This highlights that many staff are busy and may not see the value of attending these workshops, and it is not mandatory, so it is not their priority. However, they could negotiate with the administration to change the time and update the content of the workshops, but the data shows no indication of this.

This section's findings suggest the absence of effective communication channels between departments and individuals, as staff members prefer not to engage in conversations to express their perspectives. Because there is no sign of administrative

involvement, this might imply that the academic head believes everything is fine. This section implies that this school culture does not encourage enough internal communication and cooperation, which may have an impact on the implementation of proper OSCEs.

External collaboration

This section explores the external collaboration with hospitals and other medical schools that influences the assessment of this school. The collaboration with a neighbour hospital and a neighbour medical school already exists.

> *'We sometimes call outside examiners from another college or the hospital to help us conduct those OSCE stations… (However,) it is basically not well established and not well standardised, so I think this collaboration needs to be enhanced.' (Interview1-AL).*

Another aspect of collaboration is with the nearby medical school.

> *'For some of our exams, we invite outside examiners. For example, the community medicine course OSCE is one of the OSCE tests to which we invite colleagues from (medical school x) as outside assessors.' (Interview3-AS).*

The above quotes suggest that external collaboration is beneficial for their OSCE implementation. However, the OSCE designers indicate that current collaboration is inadequate.

> *'If I have more authority, I will make more contracts with some doctors from the hospitals to help teach and examine students in the OSCEs. We can also have real patients or even use hospital facilities for the exams. That could help a lot.' (FG1-D1).*

Another OSCE designer affirms that good collaboration with hospitals and other medical schools would help this school with some assessment challenges that it is facing.

> *'The lack of examiners is an issue that could be easily solved by cooperation with colleagues from (medical school x). But I need to admit that we lack this cooperation in most of the specialities, and not all staff have the initiative or interest to contact staff from outside our college.' (FG1-D2).*

The above quotes highlight the lack of a workforce and some logistical issues, which we will discuss later. However, it pointed out the earlier issue of poor communication culture that influences some staff; this altitude hinders the improvement of the OSCE implementation. On the other hand, another participant appeared pleased with the collaboration with the neighbour medical school. She highlighted an additional aspect of assistance that they received from the neighbour medical school.

> *'When the neighbour college invites me to be an examiner in their OSCE, I feel happy and immediately accept that. Because I want to learn how they run their OSCE and how they deal with different things, it is very useful for me. However, I felt they do OSCE that is similar to ours, and they still have*

some negative aspects, and we do some aspects better than them.' (FG1-D4).

It appears that external collaboration exists, but all participants indicate that it is limited and needs further development. Even though the external collaboration appeared unstructured and required improvement, all participant comments indicate that they benefited from various aspects of the current collaboration.

In summary, this subtheme highlighted the lack of effective internal communication and external collaboration. Both are institutional issues that negatively affect the assessment culture. Therefore, there are opportunities for this school to improve the communication environment to allow staff to share their experiences and improve their OSCEs (Weiner, 2009). Furthermore, the school administration has the chance to invest further in collaboration with other institutions to further support their OSCE with expertise, examiners, and logistics (Schleicher et al., 2017a; Heal et al., 2019).

4.3.1.2 The Influence of assessment regulations

According to Ndoye and Parker (2010) and (Simper et al., 2022), regulations and guidelines are significant elements in shaping assessment culture. In this subtheme, I examine the role of regulations in this medical school's OSCE implementation. It contains two subheadings: the role of external and internal regulations and dependency on individual experience.

External and internal regulations

Assessment in Saudi medical schools is regulated externally by regulatory bodies (such as the NCAAA and Ministry of Education) (ETEC, 2023b) and internally by the regulations of the given university or school. There is no oversight of whether medical schools implement these regulations, apart from scheduled accreditation visits every four years and a report prepared by the school to be submitted to the accreditation agency. Therefore, there is a chance that schools may diverge from following external regulatory bodies' regulations as there is no follow-up, especially for unaccredited schools. One assessment specialist articulated:

'But ultimately, once we meet these rules from the NCAAA or whatever bodies, we are going to get the accreditation. There is no interference from their side later on. We are trying to maintain all these rules and concepts. But they will not inquire more about the OSCE or how it is being planned, conducted, or how the results are being analysed, and so on.' (Interview1-AL).

Furthermore, they are not providing any guidance on how to conduct the OSCE.

'There are no details on their website about how to examine any specific skill or a list of which type of assessment method can be used. But they only have a generic list of skills.' (Interview2-AL).

But at the level of university regulations, the only effect on OSCE is the adoption of fixed-referenced pass score requirements.

> *'The university has general specifications and general regulations… for example, sixty per cent on any course or exam to pass them' (Interview2-AL).*

One assessment expert was pleased that most of the courses are using OSCE, but he expressed dissatisfaction with the overall implementation of OSCE when he emphasised *'somehow modified OSCE'*.

> *'Well, we run the OSCE in all clerkship courses; however, it is not the real OSCE, but a somehow modified OSCE, if you like. But it is being run in all clerkships… Well, again, you see, in reality, there are no clearly stated policies and procedures that control OSCE in the college, unfortunately.' (Interview4-AS).*

Furthermore, the participants repeatedly mentioned the absence of OSCE guidelines.

> *'We do have a guideline from the assessment unit on how to write the MCQs, but for the OSCE, we do not have any guidelines. However, if it is there, we are not aware of it.' (FG1-D2).*

When the discussion in FG2 turned to the process of designing the OSCE, the participants were debating some implementation decisions; one OSCE designer interrupted his colleagues by saying:

> *'… but there is no clear process.' (FG2-D3).*

It is apparent that some OSCE designers have no robust justification for some of their choices regarding the OSCE. For example, instead of using the blueprint to decide the number of stations, one OSCE designer looks first at the available examiners and then determines the number of stations.

> *'We are changing as per the staff's availability. It depends on how many examiners we are able to arrange. Accordingly, we arrange our station.' (FG2-D3).*

This pragmatic approach to OSCE design resulted from the absence of clear OSCE guidelines advising how many OSCE stations should be held in each OSCE exam.

> *'It is different from course to course; there is no specific standard for the number of stations; it is all up to the course organiser.' (FG1-D1).*

Moreover, there is uncertainty and inconsistency about some regulations. One was about whether students who fail the OSCE can pass the course or not.

> *'There are inconsistencies, you see. To the best of my knowledge, we have not yet reached a decision regarding that. I need to ask, actually, the senior administrators about that'. (Interview4-AS).*

There was a clear sense of frustration from this assessment specialist. His tone changed to emphasise the part of his speech that was presented. He used the term 'inconsistency'

many times during the interview. It is possible that this is a serious issue at this school. This could be an issue of communication that was discussed earlier. However, it was not clear who is *'the senior administrators'* who have the power and authority to make this decision. But the senior academic leader has a clear answer to this.

> *'There is no fail for the OSCE exams.' (Interview1-AL).*

Conversely, one of the OSCE organisers in the paediatric clerkship gave a different answer for this issue.

> *'I believe that the OSCE is more valuable than written exams in the clerkship courses; that is why we considered passing the OSCE mandatory to pass the course. Therefore, if any student fails to pass the OSCE, he or she will fail the whole course. Also, the weight of the clinical exam grades is more than that of the written exams.' (FG1-D1).*

However, there is another apparent discrepancy between educators, as their answers are disparate regarding the OSCE pass mark.

> *'I think to some extent, yeah, seventy is the passing score'. (Interview3-AS).*

This is a major issue, but he seemed to have no clear answer yet. Nonetheless, the other assessment expert said:

> *'But then, this year, we have amended the cut-off point of pass and fail, and it is now sixty per cent'. (Interview4-AS).*

The above quotes highlight the ambiguity in schools' regulations and their assessment programme: whether there is a pass or fail in the OSCE, whether the OSCE grade affects the overall course grade, and whether the pass score is sixty or seventy. This reveals inconsistency and uncertainty in such an important decision, which may be associated with the lack of clarity in some institutional regulations. According to one assessment specialist, the pass mark has changed recently. However, the other assessment specialist did not seem well-versed in pass/fail decisions, so he provided inaccurate information without a clear explanation for this discrepancy. Although changing the passing score is important, it does not appear to have been effectively communicated to assessment specialists and other faculty. This finding is consistent with the last subtheme (4.3.1.1), which is that there is a lack of effective communication within the school.

Dependency on individual experience

The ambiguity of some regulations and the lack of guidelines drive the staff working on the OSCE to depend on their personal experiences rather than evidence-based practices. The absence of established roles was identified by one academic leader as a challenge.

> *'Sometimes we have to rely on personal-level actions because there are no well-structured roles, like arranging with the examiners who are coming and who are not coming. Also, the selection criteria for the examiners are not well established.' (Interview1-AL).*

Many participants pointed out that they depend on their personal experience when it comes to designing the OSCE with insufficient knowledge, as one OSCE designer explicitly stated:

> *'We do not have a medical education background. We learn by making mistakes (laughing), or let us say, by trial and error.' (FG1-D1).*

The OSCE designers indicate that they tried to mitigate this issue by putting in a personal effort to self-educate themselves and seek advice from other colleges informally.

> *'Most of what we did was based on our personal experience. We try to read in some resources, but this is not enough, or let us say, it is not always useful.' (FG1-D1).*

> *'Actually, we ask the experienced colleagues, and they will give us their experience.' (FG2-D2).*

Also, some examiners depend on their personal experience while assessing the students, apparently without training in this role.

> *'I just depend on my experience with the OSCE exams as a student before or later as a resident… Honestly, yes, it is all my experience.' (FG3-E5).*

This situation could be because of the absence of detailed guidelines for executing the OSCE. Dependency on individual experience could also imply the absence of proper training for the staff working on the OSCE. Another examiner confirmed this inference:

> *'Indeed, we face many challenges. We are trying our best to do what we think is correct, but again, it is all from our own experience. This recalls our earlier point that we need professional training and being trained in what is considered good and what is considered bad practice in the OSCE; I always have these inquiries in mind.' (FG3-D6).*

In the fourth theme (Resources and infrastructure setup 4.3.4), I will present the findings regarding the training status of this medical school.

In summary, there are some generic external regulations, but regulatory bodies do not follow up on them regularly. The internal regulations seem inadequate to design successful OSCEs, with some inconsistency and ambiguity. As noted in the document analysis, the analysis indicates a lack of guidelines advising the faculty on how to implement the OSCE; consequently, the majority of faculty working on the OSCE rely on their personal experience. Dependence on individual experience may undermine the OSCE design and implementation of an exam for which the school cannot guarantee its quality or defend its results (Ndoye and Parker, 2010). Therefore, such an OSCE may

not meet the criteria of good assessment, such as validity, reliability, and educational impact.

4.3.1.3 The desire for improvement

This subtheme highlights the eagerness of this school staff for improvement. This is supported by being a newly established school and the participants' vision for development.

Being a new medical school

The data highlights the interviewees' views regarding being a newly established medical school.

> *'We are a new college with many junior faculty members, so we have remarked that the faculty are flexible to learn new things and are happy to use new (assessment) tools.' (Interview2-AL).*

The quote shows that this current status allows for flexibility and freedom in their assessment system. It regards being a new school and having many junior faculty as a strength point. Also, it describes the OSCE as *a 'new tool',* which may indicate the novelty of their assessment programme. These characteristics are shared among many medical schools in SA duo to the recent rapid expansion of medical schools, so they are unencumbered by history. Being a new school allows them to make changes easily compared to old schools. The female designer 1 articulated this clearly:

> *'We have flexibility in questioning and doing things differently. We are not fixed; I mean, we do not have to do things in a certain way, and that is it. We are not like what is happening in the old colleges—that is why you can see changes from course to course and from year to year. Nobody says we used to do this thing in this way, and we do not want to change, no; no one says that. All want to change for the students' benefit.' (FG1-E1).*

She emphasised the freedom to innovate and change without being constrained by old systems. Therefore, new medical schools have a good opportunity to utilise the flexibility of creating or adopting appropriate policies and regulations with minimal resistance from previous legislation (Simper et al., 2022). However, as much as this has been seen as an opportunity, it is also a challenge, as on assessment specialist suggested.

> *'We need to develop these rules and regulations, but to run a new medical college without having clearly stated policies and procedures…This is a problem, you see.' (Interview4-AS).*

In short, being a new medical school with new faculty, using a new assessment instrument is seen by most of the participants as an opportunity to implement the OSCE well. Nonetheless, the last quote highlights that this is an opportunity only when they get the foundations right, i.e., establishing clear policies and guidelines.

Vision for development

Many quotes pointed out that the interviewees have the desire and the vision to improve the OSCE. Although the assessment regulations subthemes demonstrate the absence of effective regulations, most staff are very keen to support students and improve OSCE implementation. Staff repeatedly mentioned their desire to do things right and showed their readiness to do things differently for students' benefit.

> *'Yeah. Yeah. We are still, you know, evolving or, if you like, trying to do our best to improve the OSCE in our college, and I think—due to many shortages—we did not reach the optimal type of OSCE till now. But we are working toward excellence, Insha'Allah.' (Interview3-AS).*

There is a sense of determination among some faculty to improve their OSCE. This interviewee conveyed that their implementation is currently suboptimal. In this context, the word 'Insha'Allah' can be translated as 'definitely will', reflecting their deep desire for improvement. Other participants identified some areas for OSCE improvement.

> *'If I had more authority, then I would invest in running an appropriate OSCE in our college. Because, again, all that really matters is certifying competent doctors, who have the sufficient clinical competency to practise medicine safely in an ethical and professional way.' (Interview4-AS).*

> *'I wish we had a well-prepared place and a good number of real patients. Also, I hope we will be able to (video) record the station so that we could have a reference.' (FG4-E3).*

This signifies the lack of proper facilities to deliver a better OSCE, which I elaborate on in the fourth theme (4.3.4). This may explain why some departments do not utilise the OSCE.

> *'In my department… we do not have OSCE examination for the students because, well, I do not know why; I do not know; I have to think about it… There are lots of limitations we are trying to improve.' (Interview2-AL).*

This suggests that this interviewee probably wanted to use OSCE in his course, but he quickly realised some challenges. He could not provide any explanations, but the analysis suggests that a lack of supportive regulations and resources may impede his desire (Ndoye and Parker, 2010).

In summary, the first subtheme suggests that this school's culture is characterised by ineffective internal and external communication, which renders OSCE implementation challenging. Similarly, the regulations are not clear to all faculty members, implying that they were not effectively disseminated among the staff, leading them to rely on their experiences rather than the regulations. These factors are likely to lead to suboptimal OSCEs and implementation disparities, compromising their validity and reliability. However, another institutional cultural aspect is characterised by the faculty's desire for improvement (Alsharif, 2011). As a newly established school with enthusiastic junior faculty, this school has a unique opportunity to implement a high-quality OSCE (Simper

et al., 2022). The analysis reveals that the staff is aware of their OSCE's suboptimal implementation, but they are willing to improve it; therefore, they offer suggestions.

4.3.2 Second theme—Faculty expertise and practices

This theme explores the faculty expertise and practice impact on the current OSCE implementation. It captures the reality of current OSCE practice based on the perceptions and understanding of academic and administrative staff. This theme contains two subthemes, as in Table 10.

Table 10. The subthemes of the second theme and their description.

Subthemes	Description
Faculty awareness of the OSCE	Explores faculty understanding of the OSCE and their perception of their OSCE
Faculty efforts in planning and preparation	Understands the processes that the people working on the OSCE follow to operate their OSCE

4.3.2.1 Faculty awareness of the OSCE

This subtheme reflects faculty understanding and perceptions regarding the OSCE's nature as a clinical assessment instrument. In addition, it explores faculty perceptions of the OSCE's current status and implementation at their institution.

Faculty understanding

The faculty's knowledge of OSCE's nature reflected their awareness of OSCE's strengths and limitations. The key benefits of the OSCE, as outlined in the literature (Khan et al., 2013b; Abdulghani et al., 2015; Harden et al., 2015), are well-known to many participants. They highlighted some of the criteria that make the OSCE more beneficial than other clinical assessment instruments.

> *'I really prefer this type of assessment because it is fairer, blueprint-able, and more practical than the old system of long and short cases. In the old exams, the examiners dominated the exam, and usually, that was not fair. They (examiners) ask different questions to each student, you see, but the OSCE is much fairer.' (FG4-E4).*

The aforementioned quote, among many others, demonstrates the varied reasons why faculty members prefer this instrument.

> *'It (the OSCE) reflects the understanding and psychomotor skills of students and their ability to synthesise and integrate course content.' (FG1-D5).*

An additional indication of the faculty's knowledge about the OSCE was when one assessment specialist described the three main phases needed to implement a good OSCE. These include proper planning, proper delivery, and proper post-hoc analysis.

> *'Well, OSCE—in my personal opinion actually—is a very good way for assessing the clinical competencies, especially if it is run in the proper way, ok,… it should be preceded by a planning phase, which is not an easy task, ok, and then running the OSCE itself, ok, and then the post-OSCE, so to speak, analysis. So, if these three stages are done in the right way, then the OSCE is actually a very good way of assessing clinical competencies. Otherwise, if it is not appropriately done, then it would not be a good tool for assessing clinical competencies.' (Interview4-AS).*

On the other hand, it was evident from the responses of the participants that they were aware of the OSCE's limitations.

> *'The most (prominent) one, which I consider the most important disadvantage for the OSCE, is that students are testing (the patients) in a compartmental manner rather than as a whole.' (Interview3-AS).*

Another interviewee stated that they are aware that the OSCE has limitations but that it is still a valuable tool and that these limitations should be taken into account.

> *'The OSCE reflects the students' clinical skills, but from another point, it does not reflect the reality. Because of the time, most OSCEs are five to seven minutes, during which we ask the student to take the clinical history, which, in real practice, takes much longer.' (FG3-E5).*

The data indicate that faculty members are knowledgeable and reflective regarding the OSCE as an assessment instrument. Although some participants identified weaknesses in the OSCE, they tended to view it as a valuable clinical assessment instrument. In general, the majority of participants have adequate knowledge and awareness of the OSCE's advantages and disadvantages.

Faculty opinion on their OSCE

The preceding section explores faculty perceptions of the OSCE as a clinical assessment instrument. In this section, however, I traced the opinions and perceptions of participants regarding their current OSCEs. This also looks into the impact of their OSCE on the students' future practise. According to one OSCE designer, OSCE execution in this school is different from course to course:

> *'There are many differences from course to course in the understanding and use of the concept of the OSCE method as well as calculating scores.' (FG1-D2).*

However, the application not only varied between exams but also did not meet the administration's expectations.

> *'Well, I think there are defects for sure... I think we are above average, but we did not reach the level that we were aiming for.' (Interview1-AL).*

Also, an examiner seems disappointed about the current OSCE implementation.

> *'I am not very happy; however, I think we are doing the best OSCE that we can produce based on the resources we have, I guess.' (FG4-E3).*

Participants in the quotes above suggest a lack of consistency in the design and suboptimal implementation of the OSCE at this school. The analysis suggests that this variation could be linked to the absence of clear regulations and guidelines discussed in the first theme (4.3.1.2). Another explanation is that each course has its own OSCE exam, suggesting that different designers have various degrees of expertise and approaches to designing the OSCE.

There was also a sense of frustration and criticism among some faculty, reflecting shortcomings and dissatisfaction with the current OSCE implementation. At the same time, this demonstrates their understanding of what the OSCE should look like.

> *'I am not particularly happy with the assessment of the clinical competencies in the college because there is no systematic way of assessing the clinical competencies… They claim they are conducting an OSCE, but in reality, what they conduct is not an OSCE. It is just a sort of short case assessment.' (Interview4-AS).*

Notably, the above quote indicates that they are not conducting a proper OSCE, but rather 'a sort of short case assessment'. Similarly, another participant acknowledged that their OSCEs are occasionally poorly structured and do not assess psychomotor skills.

> *'Sometimes we do not have a real OSCE exam, they are calling it OSCE, but it could be OSPE or oral exams, but it is not a real OSCE. This is also a weakness… In some stations, there are no skills have been examined; I mean, no psychomotor skills in some stations. It is just a scenario, and they are discussing (it) with the student.' (Interview2-AL).*

I asked the participants above for explanations for their opinions, and I presented their responses in the appropriate subthemes (4.3.3.24.3.3.3). For instance, some have pointed to the lack of assessment experts overseeing the OSCE implementation, and others have cited the lack of resources and manpower, both of which are discussed in the fourth theme.

Although many participants were dissatisfied with their OSCEs, many indicated that it still had a positive educational impact, demonstrating the utility of this assessment instrument.

> *'We all agree that OSCE is very useful in terms of its educational impact. Students will hardly forget cases that they have examined in the OSCE. It is a very useful tool.' (FG1-D4).*

Moreover, one interviewee affirmed the OSCE's positive impact on students' hospital performance. He cited a piece of validity evidence from hospitals where some students received their training:

> *'They actually performed well in the hospitals based on the employers' reports. So, this would mean that despite the fact that we do not run OSCE in the proper way, we still need to keep running this OSCE because it has an influence on the performance of students in the hospital.' (Interview4-AS).*

Although he acknowledged that their OSCE is suboptimal, it is still able to enhance students' clinical skills. However, the following OSCE examiner believed that what they were doing currently was insufficient. Therefore, he suggests that providing quality feedback and formative OSCEs could enhance the OSCEs' value.

> *'I believe that OSCE should have a better educational impact. It already does have some, but I mean, it could be more useful in terms of its educational value. For example, I occasionally talk to the resit students and give them feedback about their poor performance. I may also provide some hands-on training in the OSCE and explain how they could do better. It turned out to work very well, and they did pass the reset exam. However, what they have in the OSCE will stay the same during their internship, residency, fellowship, and for their whole practice.' (FG4-E1).*

In summary, this subtheme explores the faculty's awareness of the OSCE. Most participants demonstrated good knowledge and comprehension of the OSCE theoretical concept, which could have a positive impact on their assessment design (Berendonk et al., 2013; Bearman et al., 2017). They also criticised many aspects of their OSCE application, claiming it was suboptimal. Although the interviewees agreed that their OSCE has some educational impact and positive effects on students' future practice, they also acknowledged that it needs further improvement.

4.3.2.2 Faculty approaches in planning and preparation

The planning and preparation for the OSCE involve a number of steps and activities. This subtheme investigates how this school plan, prepare and write the OSCE. The section then presents some of the examiner's suggestions for a better OSCE implementation.

Preparation and writing of the stations

Boursicot (2010) and Bearman et al. (2020) emphasise the importance of an established preparatory phase for the successful implementation of an OSCE. One participant acknowledged that the OSCE is a difficult assessment to prepare for.

> *'So, it is a demanding assessment tool, ok, in terms of planning, training faculty members, the logistics, deciding the appropriate stations, finding simulated patients, writing scenarios, and so on. It is very demanding, in fact, you see'. (Interview4-AS).*

However, some OSCE designers indicated that they would start writing and preparing for the OSCE a few days ahead of the exam, as the following OSCE designer did.

> *'I worked one week ahead of the exam to write the stations, arrange for the venue, tools, mannequins, and so on.' (FG1-D2).*

Another OSCE designer learned to allow more time for the preparation phase.

> *'We usually tend to underestimate the time needed for OSCE arrangements. We are getting better from year to year as we learn from mistakes, but the*

most distinctive thing that I have learned and could make my course OSCE better is the early preparation for this exam.' (FG1-D5).

One assessment specialist expressed his dissatisfaction regarding the OSCE designers' arrangements for the exam, especially the lack of proper blueprinting.

'There is no appropriate planning phase that should precede the OSCE. For instance, there is no proper blueprinting.' (Interview4–AS).

Whereas one OSCE designer does not do any blueprinting, another does:

'We also do not have a blueprint for OSCE yet.' (FG2-D1).

'Our paediatric course is huge, so we need to blueprint all topics and prioritise what could be an OSCE station and what could be for the written exam.' (FG-D5).

This shows the disparity between courses in preparation for the OSCE, suggesting the absence of quality assurance measures. A similar issue is with writing the stations, as each course writes its own stations but without any review for the quality or appropriateness of the stations, as stated in the two examples below.

'Each of us writes a station that he has experience in its topic; I mean, write the station with its checklist. Then, the course organiser collects the stations from us.' (FG4-E1).

'In the general surgery course, we are a bit different. The course director consults all (or) most of us on the stations and the checklists. After that, he writes them.' (FG4-E3).

The following OSCE examiner picks up on the disparity in the preparation process:

'There are a lot of challenges in the standardisation process, like writing the questions in a way that cannot be misunderstood and ensuring that no students get confused by how the questions are stated.' (FG3-E3).

However, the OSCE examiner below highlighted one example of the lack of clarity in a station's scenarios, which suggests a weakness in the reviewing process for the stations before implementation.

'In the ICM (Introduction to Clinical Medicine) course, I noticed some defects in one scenario, as when the students read it, most of them get confused. And then they keep asking me for clarification; it is not well written.' (FG3-E1).

The data indicates that each course has its own way of designing and writing its OSCE without the assessment unit or examination committee supervision. The data suggest that some courses may not have allotted sufficient time for preparation and groundwork, while others may not have a blueprinting phase, both of which would undermine the OSCE's validity (Abdulghani et al., 2015; Raymond and Grande, 2019). This reflects the lack of guidelines, as there is no systematic approach to their OSCE implementation across the medical school. The analysis reveals some challenges in writing professional

stations and that the quality of the OSCE largely depends on individual (i.e., course organiser) experience and preferences. The OSCE's success would be hindered if its preparation phase was inadequate (Bearman et al., 2020).

Examiners suggestions for improvement to the OSCE

Many OSCE examiners made some suggestions for the OSCE designers. For example, they wanted to be informed about their station and briefed about the exam before commencing it.

> *'Most of the time, when I am invited to be an examiner, I do not know what the station is about. So, I find it difficult to comprehend what exactly I need to observe because the OSCE organisers provide the checklist ten minutes before the start of the exam. There is no enough time to orient myself.' (FG3-E5).*

Another examiner emphasises the advantages of this:

> *'Ten or even thirty minutes of preparation for the examiner is not enough. The examiner needs to know her station one or two days ahead, to prepare herself and discuss any ambiguity. This could enable us to perform better.' (FG3-E3).*

Furthermore, an examiner indicated the importance of briefing the examiners before the exam:

> *'We always wanted the OSCE organisers to explain the exam to make things equally clear for all examiners. Because we always have points that need clarification. But sadly, this is not normally happened.' (FG3-E5).*

Although briefing is an important step, both before the day and immediately before the OSCE, to improve the validity of examiners' assessments (Gilani et al., 2021; Malau-Aduli et al., 2023), no evidence indicates that the OSCE organisers responded to those suggestions. A possible explanation could be related to the first theme's discussion of poor communication culture, so the examiners' opinions are not effectively communicated to the right people.

In the second theme, the data suggest that faculty members generally appreciate the OSCE as an instrument; however, they acknowledge that their implementation is suboptimal. For example, the participants mentioned a short planning phase, some difficulties in writing the stations, and the need for examiners to be briefed on their stations and know them beforehand. The analysis indicates that each course approaches the OSCE differently, placing less emphasis on proper blueprints, early preparation, and station reviews. These findings suggest that the quality of the OSCE's implementation is questionable. The OSCE's quality and design issues are the focus of the following theme.

4.3.3 Third theme—The OSCE quality and design

This theme investigates the effect of current OSCE procedures on the quality and design of the OSCE. It highlights the key practises of this school's OSCE and how they influence its current implementation. This theme includes the three subthemes shown in Table 11.

Table 11. The subthemes of the third theme and their description.

Subthemes	Description
Quality issues	Explores how this medical school ensures the quality of their OSCE implementation and the issues that affect it
Theory-practice gap	Inspects the mismatch between OSCE literature and this medical school OSCE practises
Reliability issues	Investigates the factors that influence the reliability of their OSCE

4.3.3.1 Quality issues

This subtheme investigates how the faculty ensure the quality of their OSCE. It also explores the impact of the language used in the OSCE on the validity of the exam.

Challenges to quality assurance

One participant acknowledged that she did not know about the quality of the OSCE in which she participated.

> *'I do not know; honestly, I do not know whether our OSCE is good or bad; I have no idea. We need an expert to evaluate our performance.' (FG3-E3).*

Likewise, other courses rely solely on the hard work of their instructors to implement the OSCE, but no quality measures are employed.

> *'We design our OSCE in the department without support from other departments. That is why the OSCE design is different from course to course. But we do our best in preparing and conducting the OSCE, and we usually believe our exam is good enough.' (FG3-E4).*

In fact, doing their best does not imply that the OSCE is of high quality. One academic leader described the feedback received after the OSCE.

> *'There is no external feedback... But we occasionally receive feedback from the staff or even the students.' (Interview2-AL).*

However, one participant mentioned three approaches they usually use to evaluate the OSCE's overall quality.

> *'We receive feedback from students, general feedback about the course, including the assessments. We also, as examiners, discuss the exam*

> *afterwards and compare the results between the tracks to see if anything is alarming. But we are not comparing our OSCE with the OSCE of other colleges, so we do not know whether our exam is better.' (FG3-E2).*

Nonetheless, the feedback that they receive is not specific to the OSCE. With inquiry, it becomes clear that the examiners' discussions are informal and do not report any issues to the OSCE designers. In further inquiry about what are they doing with students' results? She indicated that they only look for the students' results to look for any red flags, e.g., high failure rates in some stations.

Although the psychometric analysis can provide valuable numerical indicators to evaluate the quality of the OSCE (Pell et al., 2010), this school did not utilise any quality metric for the OSCE.

> *'We are not analysing the results of the OSCEs.' (Interview4-AS).*

> *'We have psychometric analysis for other written exams, but not for the OSCE. No, not yet. We do not have this.' (Interview3-AS).*

The data shows no explicit protocol to ensure the quality of the OSCE. Nonetheless, occasional feedback from some stakeholders, general feedback about the course, and simple observation of the students' OSCE scores are performed, which could be insufficient to ensure the quality of the OSCE and assure the stakeholders.

The language as a barrier

Since all students are Arabs and all teaching and assessments are in English, a significant concern has been raised that many students struggle to communicate well in the stations.

> *'I consider the language, English as a language for communication in the OSCE stations, is limiting students' ability to express their thoughts and knowledge properly, which annoys me too much. So, they tend to speak in Arabic in their local accent even.' (FG1-D2).*

The language issue affects not only students but also SPs. Almost all SPs cannot speak Arabic or English, as most speak Urdu or Hindi.

> *'Yeah, yeah. Because of the language barrier. So, they are not interacting well with the students and cannot simulate the real case scenario correctly. So sometimes they overact.' (FG2-D5).*

This makes communication with the SPs challenging for the students, affecting the quality of communication and causing misunderstanding. Thus, language-related issues may have a negative impact on students' performance (Moeen-uz-zafar et al., 2015; Almisnid, 2016). So, this could turn into a validity and authenticity issue (Diab et al., 2019).

In summary, this school does not employ effective quality assurance measures for the OSCE, so the staff cannot determine the quality of the OSCE that they design. Some

staff members have observed that the OSCE's language negatively impacts the efficacy of communication between students and SPs. Both issues cast doubt on the effectiveness and credibility of the OSCE.

4.3.3.2 Theory-practice gap

> *'I know what the real OSCE is, and I know what we are doing. There is a gap between the two. That gap needs to be filled.' (FG2-D3).*

The above quote inspired me to compose this subtheme. It explores the mismatch between the OSCE literature and the practice of the OSCE in this school. I discovered five issues in which there are gaps between theory and practice, i.e., what the literature recommends and what this school is executed. The five headings are 'only the OSCE', 'OSCE design confusion', 'two examiners', 'unwanted examiner behaviours', and 'lack of feedback to students'.

Only the OSCE

The internal regulations at the level of this medical school emphasise the importance of using multiple performance assessment tools. However, in practise, the OSCE is the only summative instrument used in their performance assessment programme. The participants barely mentioned clinical assessments other than the OSCE.

> *'The OSCE modality of assessment is one of the major tools that we are conducting in assessing our students.' (Interview1-AL).*

Another participant declared that the other instruments are formative.

> *'I think they mainly rely on the OSCE for assessing clinical competencies, besides other tools, but the main (tool) is the OSCE… Other tools are direct observation in the hospital and portfolios, but both are formative'. (Interview3-AS).*

No single assessment instrument would be able to test all facets of clinical competencies, as each instrument has its own advantages and disadvantages (Harden, 2015; Norcini et al., 2018). Therefore, the OSCE as the only summative clinical assessment tool goes against the best practice, which is to use a variety of clinical assessment tools in the assessment programme (van der Vleuten et al., 2010; Boursicot et al., 2020)

OSCE design confusion

I noticed confusion among some staff in naming different types of assessments and the OSCE. For example, one OSCE designer counted picture-based scenarios as OSCE stations:

> *'We have the clinical assessment in two parts. One of them is face-to-face discussion, whether with a simulated patient, a real patient, or a clinical scenario, where students take the history, and we assess them on their history-taking. The other part of the OSCE assessment, as they said, is a*

> *picture-based scenario, so they have a picture with a scenario, and students have to answer it.' (FG2–D5).*

Another interviewee described the spot diagnosis, while my question was about the OSCE. She described part of the OSCE in their course and said:

> *'The other part of the OSCE is spot diagnosis; we prepare a picture, or lab results, for example, in the station, and students have to identify the clinical diagnosis and answer the questions in written.' (FG3-E3).*

Furthermore, in another course, they introduce a 'written OSCE' in the midterm exam:

> *'In our course 'psychiatry', we do have a written OSCE for the midterm, ok. It is a written OSCE case with questions about the diagnosis and management. We use this method because we do not have enough examiners or even patients. But, for the final OSCE, ok, it is a face-to-face exam from three to four stations.' (FG4-E2).*

This midterm exam looks like a written exam as no psychomotor skill has been examined; it could be another assessment type, but not OSCE. According to the quotes above, some OSCE stations do not require candidates to demonstrate clinical skills, instead requiring them to answer questions akin to applied knowledge exams. The literature does not generally support picture-based scenarios, spot diagnosis, and 'written OSCE' as an OSCE (Khan et al., 2013b; Abdulghani et al., 2015; Harden et al., 2015).

In another FG, I paused the discussion to check my understanding of what they described as an OSCE. SO, one OSCE designer apologetically explained this confusion:

> *'Sorry, I think we have different naming for different exams; we have some confusion in this regard. Some of us use OSCE, and they mean other exams.' (FG1-D2).*

I was confused in the early interviews due to this naming issue. However, when I became aware of this variation, I explained to the participants what I meant by the OSCE. I found this necessary to obtain a common ground and clarify what my questions are about. However, one OSCE designer was aware that their exam was not an OSCE, but his course still referred to it as such, and he seemed unable to change this.

> *'We do not apply OSCE in a real way; we are just doing the picture-based scenarios in slides and call it OSCE. And according to the slide, we give them scenarios and space for short answers.' (FG2-D1).*

Whereas a previous subtheme (4.3.2.1) suggested that many participants understand the OSCE concept, this subtheme indicates that many others do not. This naming variation might be considered a side issue, but it may reflect the level of uncertainty in practise and the absence of assessment fundamentals. The data suggest that students' clinical skills might be inadequately assessed because some staff members do not use the appropriate performance assessment instruments.

Two examiners

The faculty in this school has the notion that two independent examiners in each station are important. The assessment policy of the school indicates that the OSCE requires two independent examiners.

> *'Per the policy… there should be two independent evaluators, and then we take the average. This is the optimal method approved in our college.' (Interview2-AL).*

Most participants expressed their worries about the bias in the examiner's judgment and believed two examiners were the solution.

> *'We are supposed to have two examiners on the same station to decrease the chance of bias. But this is not applicable because of the lack of staff, so we do not actually do it.' (FG2-D2).*

In the surgery clerkship, they managed to recruit outside examiners and internal examiners despite the shortage of examiners (4.3.4.1).

> *'In each station, we have two examiners, one internal from the college and one external from the hospital. Then, we take the average. to overcome the overrating or under-rating.' (FG1-D3).*

In the above quote, they choose to use two independent examiners at the expense of using more stations (4.3.3.3). Having more stations would produce a more valid and reliable exam than an exam with a few stations (Boursicot et al., 2011). Moreover, the psychometric literature suggests having more stations with one examiner in each station is more important than having a few stations with two independent examiners in each (van der Vleuten, 1996; Abdulghani et al., 2015; Harden et al., 2015). Although some research in postgraduate settings suggests that two examiners (e.g., combining high- and low-stringency examiners) can produce reliable results, this method can only be used when there are enough examiners available, which is not the case at this medical school (McManus et al., 2006). This issue indicated a lack of evidence-based OSCE practices, indicating the necessity for assessment experts at this school as well as further training for faculty working on OSCE, as discussed later (4.3.4.1).

Unwanted examiner behaviours

Some examiners display undesirable behaviours during the OSCE stations. The behaviours range from unwanted facial expressions to interacting with the students and neglecting the checklist. For instance, the most common problem indicated by the participants is that many examiners are interfering with the students while performing the skills.

> *'Sometimes the examiners interfere with the students taking a history or performing a physical examination, and sometimes they convert the OSCE into, if you like, an oral exam, you see, which is against the philosophy of the*

> *OSCE... Some examiners are nodding or showing agreement; they need to be neutral' (Interview4-AS).*

This assessment specialist's statements are consistently critical and discontent with the status quo. For example, he regularly used the term *'you see'* as a form of emphasis; it may also convey that this issue happens frequently. Therefore, he described their OSCE: *'it is not the real OSCE, but a somehow modified OSCE'* as I cited earlier (4.3.1.2). Similarly, one designer elaborated on this situation:

> *'I believe the problem is not with the OSCE; the problem is with the examiners. They do not understand what they are doing. We gave them a checklist, and we expected them to just observe and confine their subjectivity. But they start a discussion or give feedback during the stations! Please keep quiet… Then the student loses his confidence. Once the examiner gives feedback to the student during the exam, like that you are doing something wrong, the student starts fumbling.' (FG2-D3).*

In the above quote, the OSCE designer was annoyed by this behaviour and its negative consequences. This reflects his awareness that this problem is not because of the OSCE but because of the examiners' behaviour. Another significant concern is described below:

> *'Sometimes, I mean, it is too common that the examiners keep asking the student during the station about questions or skills that are not mentioned in the checklist or even the focus of the station.' (FG3-E3).*

While the examiners in the above quote agreed that this was inappropriate, the examiner in the quote below implies that it is acceptable behaviour.

> *'On some points of the checklist, we should have great flexibility. For example, sometimes I tend to ask students extra questions to push them up and give them more marks.' (FG4-E3).*

If the examiners interact with the students in ways not specified in the checklist, the test's validity, reliability, and fairness will be jeopardised (Boursicot and Roberts, 2005; Reid et al., 2016). The data suggests that the examiners may require additional training that instructs them about OSCE principles before participating in the exam (Pell et al., 2008; Fuller et al., 2017).

Lack of feedback to students

In accordance with Boud (2015) and Hattie and Clarke (2018) recommendations, most of the staff believed in the importance of feedback to the students. However, they have issues preventing them from implementing this.

> *'We are not doing this although it is very important. Why are we not doing it? Because first of all, our OSCE is not the ideal OSCE exam. Number two, there is no structural feedback method to give feedback to the students. And thirdly, our students are not, I mean, trained on how to accept feedback. And there is no video recording, and this is the fourth thing. Also, we have never been taught how to give feedback.' (FG2-D5).*

Moreover, another participant expressed that they could not provide feedback due to the time restrictions.

> *'I wish I provide them with timely feedback, but there is no time in the station. I think it is important to give them immediate feedback, but how?' (FG1-D2).*

The majority of those who work for the OSCE agreed that the lack of a structured, built-in feedback mechanism is a contributing factor.

> *'We still do not give feedback to the students. I think this is one of the problems that we have to fix… We do not have, you know, a systematic way of giving feedback to a student after the test.' (Interview3-AS).*

All quotations above discussed oral feedback, but no one mentioned written feedback, which could be a plausible way to provide it (Ngim et al., 2021; Sterz et al., 2021). However, one OSCE designer, in his comment below, indicates their need for support, guidance, and training to implement the feedback on the OSCE.

> *'Because no one told us how to do so (giving feedback). Simply, that is the issue. I also do think our OSCE is not based on Medical Education evidence for assessment, such as giving feedback.' (FG2-D2).*

The data show that all participants believed that feedback was essential, but none was provided. The participants noted several reasons for this problem, including a lack of expertise in implementing it and time constraints, which are similar to the challenges faced by other medical schools (Ngim et al., 2021). According to the data, the analysis indicates their need for assessment experts to assist them in providing effective feedback to their students' OSCE.

In summary of this subtheme, this school has some issues that demonstrate a disparity between theory and practise in their OSCE implementation. These issues include the use of only one performance assessment, the confusion of other assessments with the OSCE, the use of two independent examiners, undesirable examiners' behaviours, and the lack of feedback to students. These issues are against the criteria of good assessment discussed in Chapter 1 and Chapter 2, which would undermine the utility of their OSCE (van der Vleuten and Schuwirth, 2005). The analysis indicates that OSCE staff members require effective training, clear guidelines, and a greater understanding of OSCE principles to implement high-quality OSCEs.

4.3.3.3 Reliability issues

This subtheme explores the main issues affecting the OSCE's reliability in this case study. I generated some subheadings to explore the key findings about their OSCE reliability. These include the number of stations, parallel circuits and examiners' fatigue, and marking scheme design.

Number of stations

In a previous section (4.3.2.2), one assessment expert touches on the lack of proper blueprints and the absence of an appropriate preparation phase. But here, he demonstrated one piece of evidence:

> *The interviewee: 'There is a limited number of stations because, again, some faculty members run just three stations.' (Interview4-AS).*
>
> *The interviewer: 'Sorry, how many?'*
>
> *The interviewee: 'Three stations. Can a three-station OSCE be considered an OSCE? In my opinion, no, you see. Some people go up to five; still, even five stations are not enough to be considered an OSCE. So, all these things are against the OSCE being appropriately run in the college'. (Interview4-AS).*

I was surprised by the number of stations being held for some of their OSCEs. However, there is a more alarming number. The paediatric course is a major clerkship course with only two OSCE stations.

> *'They will have two clinical (OSCE) stations on real patients, which we consider too much for students, I mean, not too much, but let us say two are enough to enable us to give good clinical judgment on the student's performance.' (FG1-D5).*

Therefore, the NCAAA commented on this issue during one of their accreditation visits to this school.

> *'The quality committee from the NCAAA ask us to increase the number of stations, as we do have a large number of topics in the general surgery course. They ask us for eight stations. We cannot provide sixteen examiners to run such an OSCE! We need more examiners and more support to do that.' (FG4-E3).*

According to the preceding interviewee, the lack of examiners is a challenge, as he believes that two independent examiners are required per station, which the previous subtheme indicated is not necessary (4.3.3.2). This suggests that one misconception would have negative consequences on other OSCE design decisions. Moreover, one OSCE designers decide the number of stations based on the availability of the examiners but not how many stations the exam should be, as indicated earlier (4.3.1.2).

Another reason for having few stations is the low number of skills taught in some courses.

> *'We teach them a very limited number of skills in the obstetrics and gynaecology course, two or three. So, students always get high marks in the OSCE because they can master the skills easily and they can easily expect which skill they will have in the OSCE.' (FG1-D4).*

However, some participants recognised that this was insufficient, so they wished for a larger number of stations, consistent with the desire for improvement discussed earlier (4.3.1.3).

> *'We want to make ten to twelve stations in a row. This is one of my hopes.' (FG2-D3).*

Furthermore, another good practise is highlighted by one participant. The OSCE designers consulted content experts' judgment to determine the duration of the exam.

> *'We depend mainly on the content experts in deciding the stations' length. For example, two to five (experts) assess how much time a student needs to respond to (each station). However, it is usually ten minutes.' (Interview3-AS).*

The following OSCE designer decided to increase the number of stations while decreasing their duration.

> *'And for us as organisers, this is fair because before it was ten minutes, but with the increasing number of students, it gets like a five-hour exam. So, we decreased it to eight minutes, which is still fair to the students. Usually, they can complete all tasks in two-thirds of the time; they can complete them within the allocated time.' (FG2–D4).*

However, as mentioned in Section 4.3.4.1, there are times when they arbitrarily determine the length of the stations because there aren't enough examiners.

The number and duration of stations is a strong indicator of OSCE reliability (Harden, 1990). The number of stations indicated by the data would be considered a small number incapable of producing reliable results (Swanson et al., 1999; Abdulghani et al., 2015; Heal et al., 2019). The findings suggest that this can mostly be due to four reasons: holding an OSCE for each course, some misconceptions about the OSCE design principles, lack of clear guidelines (4.3.1.2), lack of examiners and other logistics, as discussed later (4.3.4).

According to the data, holding an OSCE for each course would allow a limited number of skills to be blueprinted and examined, undermining the exam validity. Moreover, holding an OSCE for each course would place extra pressure on the constrained resource. However, designing a single well-structured OSCE exam by the end of the academic year would alleviate these concerns. This idea is proposed by the fact that one course stated that roughly two skills would be taught and assessed. On the other hand, history tacking could be examined repeatedly in each course OSCE, limiting the opportunity to test other skills.

Parallel circuits and examiners' challenges

Examiner fatigue is reported frequently in this case study, for example:

> *'The exam time is very long. I always get exhausted; do you understand what I mean? We may need two tracks in all exams, but we do not have enough examiners.' (FG3-E6).*

Therefore, some courses design their OSCE to run two parallel tracks in which each has the same type and number of stations to reduce the examiners' fatigue and total exam time.

> *'Sometimes the examiners' number is barely enough for one track, but we tend to make two tracks of four stations to avoid the issues with the patients getting tired from waiting for a long time, the students in the quarantine, or even the examiners' exhaustion. In this case, we need more examiners, which, you know, is always a challenge.' (FG4-E1).*

Notably, they have hardly been able to recruit eight examiners for the two tracks, so they appear to have preferred this arrangement over having eight different stations on one track. This goes against the literature recommendations for more stations (Khan et al., 2013a). However, staff and students became worried about the variation in students' results between the tracks. As an endeavour to minimise the disparity between the circuits, the paediatric course OSCE designer did the following technique:

> *'For the paediatric course, we have OSCE in the midterm and final, keep the same examiner on the same track, and shift students between the two tracks. So, we are more relaxed after this manoeuvre, and students feel fairer than before.' (FG2-D4).*

One participant confirmed that differences in marking stringency are evident among many examiners.

> *'There are huge differences in terms of the strictness between the examiners. Some of them are too lenient, and they can reward some unacceptable performances. You can notice this from the grades of their stations… You can solve this issue with new staff, as training can be useful for them.' (FG1-D4).*

Another participant commented on her colleague statement above and said:

> *'Some of them are not fit to be an examiner in the first place; even training might not help.' (FG1-D5).*

She realised training for the examiners is not a definitive solution (Yeates et al., 2015a). However, the literature suggests that it could minimise the variation if the examiners receive similar instructions (Schleicher et al., 2017b), as highlighted later (4.3.4.1). Furthermore, another concern that this school face is with the outside examiners, as they have different expectations from students. The OSCE's reliability, i.e., inter-rater reliability, may be compromised by disparities in assessing attitudes.

> *'I can see the differences between the internal examiners and outside examiners. The outside examiners have no clear idea about the OSCE concept. Also, they do not have an idea about students' levels; therefore, their expectations are usually high. But the internal examiners are more familiar with the OSCE style.' (FG4-E6).*

While the school and the student might benefit from the outside examiners, they may be unaware of the students' skill level or the OSCE examination procedure and may

therefore score students differently. Parallel circuits and examiners' variation might pose reliability and fairness concerns as different examiners evaluate different students on the same station (Fuller et al., 2017). Therefore, OSCE designers need to employ a variety of techniques to ensure the comparability and fairness of the parallel circuits, including faculty training (Pell et al., 2008) and appropriate post hoc analysis (Pell et al., 2010; Tavakol and Dennick, 2011b).

Marking schemes design

This section concerns the design of the marking schemes used in the OSCEs of this medical school. The marking scheme is a crucial component of any OSCE station (Ilgen et al., 2015). It could be a checklist with dichotomous ratings or a rubric with a Likert scale, as both are used here.

> *'The rubric that the course organiser prepared has a range from one to three scale, like excellent, medium, and poor. Sometimes it could be a checklist which states done or not done for each skill.' (FG4-E5).*

At least one course designed its checklist using differentially weighted items.

> *'In the general surgery course, our checklist has different items, and each item has a different mark; some points have two marks, and some have one mark, depending on the importance of the point. So, it is not a done/not done checklist; it is a mark-based checklist; we found this to be more accurate and fairer.' (FG4-E3).*

However, an examiner determined that some marking schemes are poorly designed, which could result in a significant degree of disparity between examiners.

> *'They gave me the checklist a few minutes before the station started, I went through it, but I found it not appropriate. But what can I do? It is too late to discuss it with the designer. I also notice students perform poorly as the instructions are not clear enough… When I discussed this with other examiners, we agreed that the checklist was the source of this variation. It is not well built, it does not have clear items, and it is not specific.' (FG3-E6).*

This could be related to the earlier finding that there is no briefing before the exam and suggestions by some examiners to receive their stations' plenty of time before the exam begins (4.3.2.2). Things can be fixed when the above steps are followed.

> *'I remember when I received the checklist and read it before the station, I discovered an inappropriate item, so I discussed this immediately with the OSCE organiser, and we agreed to modify the checklist at the last minute.' (FG3-E2).*

One OSCE designer emphasised the importance of developing a good scoring sheet to reduce examiner variation.

> *'If we could have a well-built rubric or checklist, it would minimise the bias or the variation between the benign and malignant assessors, so to speak.' (FG1-D1).*

She implied that this *'well-built rubric or checklist'* does not exist yet. In contrast, other OSCE designers appeared pleased, believing that they had created a high-quality OSCE checklist in his course.

> *'On our course (surgery), we have no big differences between the examiners, and I think our checklist has to help us to minimise the differences.' (FG1-D2).*

Another observation by one examiner indicated that they do have some theoretical questions in the checklists.

> *'Our checklist compound of clinical or practical questions and also theoretical questions.' (FG4-E3).*

The inclusion of theoretical questions in the OSCE would be unfeasible because they could be easily and cheaply assessed in written exams.

According to the data, the participants are aware that the design of the marking scheme is paramount for helping ensure OSCE reliability. While quotes show they have different marking schemes, the data suggest that they are not always appropriately designed. Therefore, many participants indicated the need to construct clear and specific marking schemes to minimise variation between examiners and enhance OSCE reliability (Abdulghani et al., 2015).

To summarise the findings of the third theme, the three subthemes discussed revolve around one concept: OSCE quality and design. The data indicated that this medical school lacks quality measures for the OSCE implementation, as there is no pre-exam review or post-exam analysis, leaving the faculty unaware of the OSCE's quality and unable to defend its results. Additionally, the analysis revealed a gap between the OSCE practises of this school and the criteria of good assessment. For instance, they do not use any other performance assessment besides the OSCE, they tend to assign two examiners at each station, and they do not provide students with feedback following the OSCE. Further findings indicate that their OSCE reliability is compromised by a number of factors, such as conducting a small number of stations and parallel circuits without sufficient evidence of standardisation between the circuits and examiner training. Lastly, the data suggest that the marking schemes are not always of high quality, which could compromise the validity of student scores. These findings suggest that their OSCEs are implemented sub-optimally and require significant enhancements in terms of their validity, reliability, and fairness.

4.3.4 Fourth theme—Resources and infrastructure setup

In this subtheme, I explore the impact of these medical school resources on their OSCE implementation. Since the OSCE is known to be a resource-intensive exam (Gupta et al., 2010; Patricio et al., 2013), this theme investigates the relationship between the

resources available and their impact on actual OSCE implementation. This theme classified the resources into two categories, as shown in Table 12.

Table 12. The subthemes of the fourth theme and their description.

Subthemes	Description
Human resources	Examines the impact of the OSCE workforce (faculty and patients) availability and training on the OSCE implementation
Instructional and logistical resources	Explores the effect of educational activities and physical resources on the OSCE implementation

4.3.4.1 Human resources

Many personnel (faculty and SPs) work together to create the OSCE. Therefore, this subtheme investigates how each personnel's availability and training affect the implementation of OSCE.

Lack of workforce

In some of the previous subthemes (4.3.1.1 and 4.3.3.3), many participants indicated a lack of staff working on the OSCE while discussing something else. However, the lack of faculty seemed to be a significant issue.

> *'The main challenge in my department is the shortage of staff.' (FG3-E6).*

An OSCE designer described the depth of this issue:

> *'I am the course director, I am the teacher, I am the OSCE organiser, and I am the OSCE examiner! I need more staff with me during the course and the OSCE exam. So, I have to invite other colleagues to help me. So, this determines when the exam will be or how long it will be.' (FG1-D5).*

As mentioned earlier (4.3.1.2), the quote above noted that the number of staff available decides the length of the exam, i.e., the number and duration of the stations, which may mean that these decisions are unplanned. A different OSCE designer specified the lack of station writers and examiners:

> *'I need someone to help me construct the OSCE and write the stations. We need variation in the cases and stations. But now I am forced to plan for the exam and write all the stations by myself. I consider this a negative aspect of our OSCE. But the big problem is allocating enough examiners.' (FG1-D4).*

They try to overcome the shortage of examiners by inviting outside examiners, as indicated earlier (4.3.3.3).

> *'Some of the outside examiners who come from the hospital to support us in the exam; they think that students are supposed to know the details, but they only know the general and theoretical part of it. In that way, they reflect this expectation in their harsh marking.' (FG2-D2).*

Sometimes outside examiners are not academic personnel but clinical doctors from hospitals. These outside examiners have different expectations of the students; they are unaware of the students' levels, what they should know, and what they may not know. This would affect the validity of their scores, suggesting that they need training to adjust their expectations (Abdulghani et al., 2015).

Another important aspect many OSCE designers emphasise is the lack of assessment experts to support their OSCE implementation process and making decisions.

> *'Although we have a lack of staff and examiners in our course, it is still a solvable issue. The main problem I face right now is that no one is helping me design the exam, write the stations, or organise everything. Similarly, we need support to conduct the exams and to know what to do when we face a problem or when there are multiple options to do things. We need someone who knows what to do and why he or she chooses one thing over another…' (FG1-D4).*

> *'We need instructors and experts who have a good background in how to do things (in assessment).' (FG1-D1).*

On the other hand, the participants highlighted the lack of patients and SPs. Although they are essential components in many OSCE stations, some difficulties in recruiting them may pose challenges for OSCE implementation.

> *'Also, I think SPs are another limitation as there are some challenges in recruiting and training enough of them.' (Interview 1-AL).*

In some courses, the problem is more apparent. They are not using the OSCE for many reasons, but the lack of patients is one of them:

> *'Because of the lack of the staff and, you know, lack of the of patients, it does need real patients, or simulated patients, which is not available in our college. So, based on our restrictions, we have to have a picture-based exam. And we call it spot diagnoses.' (FG2-D1).*

According to the data, this medical school is having difficulty recruiting enough personnel to work with the OSCE. The analysis indicates a lack of OSCE designers, writers, examiners, assessment experts, SPs, and patients. In this situation, the OSCE's utility and feasibility in this context would be called into question. However, the data suggests they are able to conduct some OSCEs, but they may be of limited validity and reliability.

Training for stakeholders

This section explores the training status of OSCE personnel at this medical school. Proper training is crucial to ensure that all participants comprehend their roles, as this influences how the OSCE is implemented (Reid et al., 2016; Schüttpelz-Brauns et al.,

2019). The need for OSCE training is one of the most frequently mentioned aspects by all participants.

> *'I think what we are lacking is training. I think this is one of the limitations here, so we need to train more faculty members, more assessors, and more SPs. So, we need more workshops and seminars on how to prepare and conduct the OSCE.' (Interview3-AS).*

The following participant described his experience as an OSCE examiner in other courses.

> *'There is no training here for the faculty about how to introduce or conduct an OSCE exam in the proper way… Actually, there is no training, but they will ask you to come in the morning, then will assign you to one of the stations, give you the checklist, and ask you to examine the students with no previous preparations. They may brief you minutes before the exam if there is at all… So, I personally depend on my personal experience.' (FG2-D1).*

Therefore, the vision of one academic leader is to develop the faculty in this regard, which supports the desire for improvement discussed earlier (4.3.1.3).

> *'I will try to bring experts and conduct courses by national and international trainers on how to perform the OSCE examination for all of our staff.' (Interview2-AL).*

However, one senior staff member stated that she had received some training in the past.

> *'For me, I attended one workshop a long time ago. It was an optional session. But from my personal experience, the workshop was difficult to comprehend, difficult to understand what they exactly wanted, and very long. It was just lecturing more than a workshop. I, frankly, felt it was not useful.' (FG1-D3).*

This may appear to contradict other assertions about a lack of training. This can be explained by the fact that senior staff may have received training in the past, but junior staff did not. In fact, the majority of the academic staff at this school was recently appointed. They are junior academics with little experience assessing students, having mostly come from clinical backgrounds. As a result, they embarked on the OSCE without a chance to attend any training workshops since no new training sessions have been held.

> *'We frequently receive young and new faculty members. You see, we do not actually run training workshops for those newly joined faculty.' (Interview4-AS).*

This status quo highlighted the need for a structured and continuing faculty development programme. This would benefit not only new staff members but also those who have previously attended workshops to refresh and update their knowledge. Nonetheless, one OSCE designer indicated that she provided a briefing for the examiners prior to the exams, which she deemed valuable.

> *'From my own experience, whenever the checklist is accurate, and I myself sit with the examiners to give them an orientation about the checklist, then they will have less subjectivity in marking. I have noticed this clearly in different course exams. The more I show them what we are focusing on in the exam, the more homogenous scoring they produce.' (FG1-D5).*

Although briefing is considered important (Malau-Aduli et al., 2023), other OSCE designers did not provide any briefing for the examiners because they assumed the examiners already knew how to act on the stations.

> *'In our course (obstetrics and gynaecology), we are similar to the general surgery people. We are not providing any instructions to the examiners because we assume they all have the same background and know what they should do.' (FG1-D4).*

Another crucial component of training is the training of patients and SPs.

> *'There is a lack in training for the staff and SPs as well.' (FG2-D4).*

According to the data, many participants are concerned about the performance of the patients and SPs in the OSCE.

> *'Sometimes, what happens when few students examine the same patient (in the OSCE)? Then by reflex, the patient moves the body accordingly. And this happens! They also tend to show the signs that the students missed, then by reflexes, the SPs tend to help.' (FG2-D3).*

This raises concerns about the fairness and standardisation of the OSCE (Cleland et al., 2009). Another examiner conveyed the same concern and linked it to the examiners' need to interact with students to mitigate this issue.

> *'We struggle with the lack of trained SPs. We really want well-trained SPs; I think most medical colleges suffer from the same issue. That is why we need to talk with the students and SPs during the stations; we tried to, how I can say it, neutralise this deficit.' (FG4-E3).*

As a result, the lack of SP training led to a suboptimal practice in which the examiners interfered with the SP role and student during the station (Boursicot and Roberts, 2005). However, only one OSCE designer provides training for the SPs, and she is the only one who briefs the examiners, as stated in one of the above quotations.

> *'I give training to the SPs one hour prior to the exam and explain to them how they should act in the exam. I give the training to the SPs together to ensure that they receive the same instructions.' (FG1-D5).*

Her effort appears to be the exception in this medical school, which supports previous findings of a lack of regulations and quality assurance measures to ensure OSCE implementation (4.3.1.2 and 4.3.3.1).

In summary, the findings indicate a clear lack of OSCE staff, patients, and SPs, leading to challenges like a low number of OSCE stations (4.3.3.3). Furthermore, the data suggests that while a few OSCE staff members may have received training, the vast

majority did not, so they rely on their own experience, as indicated earlier (4.3.1.2). Likewise, SP training is insufficient. These issues would undermine the standardisation and affect the reliability, fairness, and quality of the OSCE results (Collins and Harden, 1998; Cleland et al., 2009; Fuller et al., 2017). The findings indicate that training for various stakeholders is one of the improvements required to improve their OSCE implementation.

4.3.4.2 Instructional and logistical resources

This subtheme explores the availability of educational and logistical resources that can facilitate the OSCE's implementation at this medical institution. This would assist in evaluating the suitability of these resources and their impact on OSCE implementation in this context.

Educational resources

In this section, I highlight the educational activities that support using the OSCE and allow students to perform well on this assessment instrument. One assessment specialist identified that their curriculum is supportive of the use of the OSCE.

> *'I would say that this is a real strength in our curriculum. It has an effect on how we use OSCE in our college. It is well designed with clearly stated courses and intended learning outcomes that align with teaching and learning methods and assessment strategies. So, we mainly prepare OSCE stations to cover learning outcomes pertaining to psychomotor and affective domains.' (Interview3-AS).*

Another staff member affirmed that they provide good teaching that can assist students with their clinical exams.

> *'I think our teaching, especially bedside teaching and the discussion, are very supportive for students. I believe this helps them a lot in the OSCE. We also provide very clear ILOs that help students prepare better for the exams with no surprising stations.' (FG3-E1).*

Furthermore, before the OSCEs, some faculty members provided more practice and teaching.

> *'We always check before the exams whether they (the students) are struggling or having problems with any topic in clinical examinations, so we can arrange for a review.' (FG4-E3).*

While this could boost students' learning and their performance in the OSCEs, some responses noted the lack of orientation for students to the OSCE exam style. Two OSCE designers realised students need to be familiarised with the OSCE as an instrument.

> *'I agree with Dr (…) when he said there is no training for the students on this tool. I think we are supposed to give them more training on this technique before they embark on it.' (FG2-D5).*

An examiner emphasised the importance of educating students on how they should manage their time during the stations:

> *'That is why I think we need to train students on how to be systematic during the OSCE station and how they can control their time to complete the task during the allocated time.' (FG4-E1).*

This suggests the importance of training students to understand the new assessment method and prepare them for the actual exam, which can be achieved by conducting formative OSCEs.

> *'What I believe is that before the summative OSCE, there should be a formative, or let us say, a demo OSCE. We should show students a sample of it. There should be a rehearsal, which should be ungraded.' (FG2-D3).*

Another OSCE designer promoted using the formative OSCE for a different reason.

> *'We do not use OSCE as a formative tool. We just use it as a summative tool to assess students' skills… Therefore, I think we need OSCE in the middle of the course with no grades; ok, this will be an educational one for them. Because at the end of the course OSCEs, the student will be grade-oriented.' (FG2-D1).*

As quotations pointed out, formative OSCE can help the students as an educational tool to support learning, provide them with feedback, and familiarise them with the OSCE as an assessment format, which is supported by the recommendation of Pugh et al. (2018). However, the formative OSCE is not utilised, so neither the formative nor summative OSCEs would provide students with feedback on their clinical skills, as indicated earlier (4.3.3.2).

In summary, the data indicated two pieces of validity evidence supporting the OSCE's use in this school, as the curriculum ILOs and educational activities align with this type of assessment (Hodges, 2003b). However, the analysis suggests a lack of specific training for students on this assessment instrument, which may undermine the acceptability of the OSCE among them. Therefore, the participants indicated the need for formative assessment to provide this training and enhance the educational impact of the OSCE.

Logistical resources

This subtheme investigates the physical infrastructure and resources available that affect the OSCE implementation in this school. These include the venue, funds, teaching hospital, and procedural support. The OSCE is a resource-intensive examination that necessitates an abundance of resources (Rushforth, 2007; Patricio et al., 2013), which are unattainable in this medical school setting.

> *'Of course, we are not happy. It could be better, you know, to have a good way to assess the students, like using the OSCE, that is important. But lack of facilities and lack of faculty are holding us back.' (FG2-D1).*

Regarding this, all participants held the same view as the participant positioned above them. For instance, they emphasised the lack of a suitable venue.

> *'I just want to emphasise that the proper place for the OSCE is very, very important. And we lack the proper place.' (FG2-D4).*

One of the consequences of insufficient space is an additional case for limiting the number of stations, as previously addressed (4.3.3.3).

> *'We are thinking of increasing the stations' number, but we find it difficult because of the limited place and examiners' number. Anyway, we still think the station number is reasonable.' (FG4-E3).*

Another participant considered that a well-established simulation centre would provide an appropriate environment for the OSCE.

> *'If we do have a simulation centre, I mean a well-prepared simulation centre. Then we can overcome many difficulties, especially the fact that we do not have an educational hospital.' (FG3-E2).*

However, there is insufficient financial support to run the OSCE and the new simulation centre.

> *'I would work more on improving the setting itself. The simulation centre still needs more rooms. We have the mannequins and some other (equipment), but we still need some more funds. I think we need more support from the university, to be honest.' (Interview3-AS).*

In response to a question about financial sufficiency, one interviewee stated that things are fine "to some extent." I tried to probe the interviewee, but his response remained consistent:

> *'(…) To some extent. To some extent.' (Interview4-AS).*

There was a brief pause, then a slow response. I believe he was uncomfortable here because he usually speaks a lot, but he suddenly overcame this. I had the impression that some things went unspoken; the organisational culture here may have prevented him from discussing this openly. Nonetheless, one academic leader described some funding challenges to appropriate OSCE implementation.

> *'Yes, yes, yes, especially at this time, the university is trying to control the financial expenses. But we need some logistics that need funds… For example, we do not have a university hospital, so we do not have authority over the hospital (administration), the patients, or the departments in the hospital. We are trying to do our own simulation centre to use it for clinical assessment'. (Interview2-AL).*

The absence of a teaching hospital impacted both the teaching and the ability to conduct the OSCE in the hospital or have enough real patients for the OSCEs. Additionally, it makes the need for funds to operate the simulation centre more pressing.

> *'As long as we do not have a university hospital and we have a limited number of patients, we need to have specialised models and mannequins to run the exam, which are not always available as well. It costs a lot of money. Also, the university hospital will provide a real environment and enough rooms to conduct enough OSCE stations.' (FG2-D1).*

Aside from logistics, the need for additional administrative support to facilitate OSCE implementation was raised.

> *'The other point is that OSCE preparation needs administrative support, not only educational work; I mean, you need them to provide funds to compensate the SPs. Say the same for improving other doable logistics. For example, establishing a bank for the OSCE stations. I think some more administrative effort could help us on this.' (FG4-E5).*

In the preceding quote, the examiner implied that administrative efforts might not be sufficient. He suggested that additional administrative efforts could alleviate some of the challenges mentioned above.

According to the data, this school does not have an appropriate place to conduct the OSCE. Although they have a building for a simulation centre, it is not functional yet. Because this medical school does not have a teaching hospital, a functional operative simulation centre could be particularly helpful. Due to the lack of a teaching hospital, the school was unable to use real patients or hospital facilities to conduct the OSCE. The analysis suggests that all these limitations can be attributable to the financial constraints that impede a proper OSCE implementation.

In summary, the fourth theme indicates deficiencies in all OSCE personnel and their training. This raises concerns regarding the school's ability to conduct valid and reliable OSCEs. Despite this, the findings show that the educational activity is appropriate and supportive for OSCE use, despite a lack of formative OSCE exercises. The findings suggest that financial considerations are significant logistical constraints that make it difficult to secure enough space to run enough stations simultaneously and operate the simulation centre. Therefore, this theme suggests that the OSCE implementation in this school would not improve unless they invest in human and logistical resources.

4.4 Case study A findings interpretation

This section synthesises and seeks to interpret the main findings from the documents, interviews, and FGs utilised in the case study. It offers a comment on the context of this medical institution. Then, I reflect on key findings that I have classified as opportunities and challenges encountered by this school during the OSCE implementation. I conclude with an overarching theme that synthesises the concepts derived from this data analysis into a central message that can be derived from this case study.

4.4.1 Notes on the context

This newly established public medical school is subject to the affiliated public university and Ministry of Education regulations in SA. The data suggests that it is bound to these institutions' hierarchy in decision-making and general regulations. Most of its faculty are newly appointed clinical doctors with limited experience in medical education. There is no current national or international accreditation for this medical school. According to the analysis, these factors influence the institutional and assessment culture, faculty expertise and practises, OSCE quality and design, and resources and infrastructure, all of which influence how the OSCE is implemented.

4.4.2 Opportunities and challenges

This section looks at the main opportunities for this medical school to improve its OSCE implementation. Besides, it highlights the key obstacles they face that prevent them from implementing better OSCEs.

4.4.2.1 Opportunities

The data analysis reveals several opportunities for this medical school to implement a high-quality OSCE. One of the most recurring themes in the data is everyone's desire to improve. This is an excellent opportunity for this school's leaders to build upon, as all interviewees share their vision to enhance their OSCE practises. For instance, many hope to introduce more stations, receive training, provide students with feedback, and allocate sufficient places for the OSCE. Although the majority of faculty members are junior clinicians, they show a good level of understanding of the basics of the OSCE. For example, the findings indicate that they valued the OSCE as an assessment tool and were able to critique their current OSCE practices. The data suggests that if they are enabled by supportive administration through providing more logistics and training, they can implement highly successful OSCEs. Therefore, knowledgeable and enthusiastic young faculty members are a valuable asset to this medical school.

The fact that this is a newly established medical school provides another opportunity to support the faculty's desire for improvement. This suggests that they are not bound by history, allowing them to be innovative and adaptable in their OSCE implementation. For example, this school would be more flexible in amending and developing its documents and regulations. This school document analysis discloses comprehensive information about the basic assessment criteria, yet no specific information guides the OSCE implementation. Therefore, they could create these documents with sufficient detail to guide the OSCE and align them with evidence-based practices. For instance, they may write an OSCE guide adopting new rules and instructions based on the best practices in assessment literature, like adopting a criterion-based standard setting approach instead of a fixed pass mark.

The findings revealed that the educational resources, i.e., curriculum and teaching activities, were consistent with and supportive of OSCE use. Moreover, the data shows that this school has a building intended to be a simulation centre, which, when completed, will be a valuable addition that helps the faculty conduct more effective OSCEs. Finally, this school has a great opportunity to obtain national and international accreditation when it takes advantage of the aforementioned opportunities and addresses the challenges discussed below.

4.4.3 Challenges

According to the findings, this school culture is characterised by ineffective internal communication between the staff members and poor collaboration with other institutions. This makes sharing expertise between the faculty members and between this school and other institutions challenging, which, according to the data, negatively impacts the OSCE exam. Similarly, the findings suggest a lack of clarity in schools' regulations and assessment programmes, posing a challenge for OSCE designers and course organisers when they make decisions about OSCE and course assessments. Therefore, the faculty is inadvertently pushed to rely on their experience rather than written regulations to make some assessment decisions.

Another challenge highlighted by the data is the lack of quality assurance measures prior to, during, and after the OSCE implementation. The findings indicate that the OSCE examination is not reviewed by the assessment unit or examination committee before delivery. Moreover, some data point to the absence of blueprinting, while others indicate a short planning phase. An obvious challenge to the OSCE's validity is the language of communication between students and SPs, as the majority of SPs lack adequate English proficiency to communicate effectively with students. After the OSCE, there is no post-OSCE psychometric analysis to calculate quality metrics. Therefore, the findings suggest that the faculty is unable to guarantee the quality of its OSCEs.

An evident challenge is the lack of OSCE personnel, including OSCE designers, station writers, examiners, and SPs. Furthermore, the findings indicate a theory-practice gap between what is considered best practices in the OSCE literature and this school practices. For example, there is an absence of feedback for students and no performance assessment instrument used besides the OSCE. Furthermore, there is a lack of training for the OSCE examiners and SPs. The data indicate that these challenges result from the lack of assessment experts. Therefore, the findings indicate the need for assessment experts to train different OSCE personnel and supervise the OSCE implementation. Finally, the data indicate a lack of a proper place and sufficient funds. The analysis suggests that a lack of human and logistical resources impeded the implementation of effective OSCE stations.

4.4.4 Overarching theme: A series of dilemmas and compromises

This overarching theme provides a clustering concept for the findings analysed in this case study, shedding light on the relationships between concepts derived from data analysis. This can comprehend the complexity of the OSCE's operational and educational processes, which involve numerous stakeholders and variables. Reflecting on the entire data set reveals that the implementation of the OSCE in this medical school is subject to several design compromises due to factors such as examiners, SPs, training, funding, and space, which forces the faculty to make difficult decisions.

The faculty voiced understanding for some of these compromises and tensions, but it appears that they are not always aware of all of them. Although the documents emphasise the need for high-quality assessments that follow the criteria of good assessments, the in-depth examination of the interviews and FGs reveals that the school is unable to conduct an OSCE that meets all the standards outlined in the documentation. The faculty makes compromises all the time without necessarily recognising them, but they are all ultimately interconnected. This reveals a quandary at each level; although participants are familiar with the majority of OSCE's good practices, they encounter challenges in their contexts that lead to these compromises.

According to the findings, this school OSCE requires substantial improvement to meet the criteria of good assessment outlined in the documentation analysis. Therefore, the data indicate that this school is unlikely to be able to do so unless it improves its assessment culture by recruiting assessment experts and investing in faculty development and resources. Moreover, to implement high-quality OSCEs, the faculty should consider redesigning the OSCE within the assessment system to promote innovation, capitalise on opportunities, and address challenges.

4.5 Summary

This chapter presented the findings with analysis from case study A documents, interviews, and FGs. Using Codebook analysis, the findings of the school's assessment documents revealed that they cover most of the components of good assessment described in the literature but with little theoretical description and some omissions in places. On the other hand, RTA was used to induce the findings from the interviews and FGs, which generated perhaps more critical results. I constructed the findings into four key themes: institutional and assessment culture, faculty expertise and practices, OSCE quality and design, and resource and infrastructure setup. The analysis and interpretations demonstrated that the current OSCE implementation at this medical school is suboptimal and highlighted the factors influencing it. The following chapter will present case study B findings and analysis in a similar structure to this chapter.

Chapter 5 Findings and Analysis—Case Study B

5.1 Introduction

This research aims to answer two questions: how is the OSCE being implemented in Saudi medical schools, and what are the opportunities and challenges offered by adopting it? Therefore, I collected data from two case studies (A and B) from two medical schools in SA to answer them. Because each school has a unique context, as discussed in the Methodology and Methods Chapter, the results of each case study will be presented in their own chapter. In the previous chapter, I presented the findings from case study A, which is an example of the public and common type of medical schools in SA (as indicated in Section 3.3.1.1).

In this chapter, I will present the findings of the second case study—case study B. Case study B is an example of, and shares the context with, the growing type of medical schools in SA, as discussed in the Methodology Chapter. The findings are derived from the analysis of twenty assessment documents, three interviews, and four FGs. The three interviews were with an academic leader and two assessment specialists, and the four FGs were with OSCE designers, writers, and examiners; Table 13 details the data sets.

As with Case Study A, I will present the document analysis results using the codebook analysis approach (Mihas, 2019; Roberts et al., 2019); second, I will analyse the interviews and FGs. The analysis of the interviews and FGs allows me to structure the findings into four overarching themes, each with some subthemes. Using the Reflexive Thematic Analysis (RTA) approach (Braun and Clarke, 2006; Braun and Clarke, 2022), themes and subthemes were generated after deep engagement with the data sets. More information on the analytical approaches used can be found in Methodology Chapter (Section 3.8).

Table 13. Details of the documents, interviews and FGs conducted for case study B.

#	Data set and participants pseudonym		Participant's description	Length
1	**OSCE documents**	20	See Table 14 for more details	64 pages
2	**Interview 1** (Interview1–AL)	1	Academic leader (AL) Five years in academic service	37 minutes
3	**Interview 2** (Interview2–AS)	1	Assessment specialist (AS) 1 Eleven years in academic service	65 minutes
4	**Interview 3** (Interview3–AS)	1	Assessment specialist (AS) 2 Thirteen years in academic service	64 minutes
5	**FG 1–Designers** (FG1–D)	6	Six OSCE designers and station writers (From 5 different courses: 1. Internal Medicine, 2. General Surgery, 3. and 4. Obstetrics and Gynaecology, and 5 and 6. Clinical Skill course)	88 minutes
6	**FG 2– Designers** (FG2–D)	5	Five OSCE designers and station writers (From 5 different courses: 1. Internal Medicine, 2. General Surgery, 3. Family Medicine, 4. Obstetrics and Gynaecology, and 5. Emergency Medicine)	88 minutes
7	**FG 3–Examiners** (FG3–E)	4	Four female OSCE examiners (From 4 different courses: 1. Emergency Medicine, 2. Internal Medicine, 3. Clinical Skills course, and 4. Clinical Skills course)	62 minutes
8	**FG 4–Examiners** (FG4–E)	5	Five male OSCE examiners (From 4 different courses: 1. Internal Medicine, 2. Orthopaedic, 3. Ophthalmology, 4 and 5. Clinical Skills course)	74 minutes
Total 8 data sets		**23 participants and 20 documents**		**478 minutes**

5.2 Documentary findings and analysis

This section is a documentary analysis of the available documents from this medical school, as discussed in Methodology Chapter (3.8.1).

5.2.1 Findings of the documents

Continuing the approach taken in the previous case study, I used the codebook analysis approach to analyse the documentary evidence from School B (Mihas, 2019; Roberts et al., 2019; Braun and Clarke, 2020). I collected twenty documents that I classified into three categories: 1) exam policies—two documents; 2) templates and forms used to prepare and design the OSCE—eleven documents; and 3) feedback forms and post-exam analysis reports—seven documents. As displayed in Table 14, I conducted a comprehensive review of the documents and summarised the relevant content in each document, along with a commentary on the contents of each. I was unable to append all documents because they are lengthy, and doing so would reveal the identity of the medical school. Therefore, I chose to provide examples of documents that will not reveal their identity (Appendices 21, 22, and 23).

Then I performed the codebook analysis, as shown in Table 15. I compared the relevant content of these documents with the assessment documents and guidance from three international organisations concerning assessment implementation in medical schools. The three organisations that I used their criteria and frameworks are the Ottawa Conference, AMEE ASPIRE-to-excellence award in assessment, and WFME, more details are in Methodology Chapter (3.8.1) (Boursicot et al., 2011; WFME, 2015; Boursicot et al., 2020; WFME, 2020; AMEE, 2023).

Table 14. This table provides an overview of the case study B assessment documents and each document's relevant content with a description.

Doc. No.	Document title	Pages	Relevant content	Description
1 & 2	Exams policies (A portion of these documents are listed in Appendix 21)	26 in total	Guidance on the school's assessment procedures. It is a procedural guide for the assessment process in this school; the relevant content includes content about: • Exam Analysis Policy. • Pre-exam Analysis. • Post-exam Analysis. • Replying to Students' Exam Feedback. • Assessment Unit Responsibility. • Courses mark distribution. • A policy of moderation of assessment.	In the two exam policies documents, this medical school generally focuses on the analysis process for the exams. It considers the steps that need to be followed to ensure a high-quality assessment. It detailed the steps and regulations for the pre-exam analysis. Many points were indicated, e.g., the exam questions should be ready and uploaded in their assessment system three days before any exam, and the need for alignment between the exam questions and the blueprint ILOs. The exams policies documents also focus on the post-exam analysis regulations. For instance, the post-exam analysis meetings should be held within five days after the exam, the students' feedback session should be after the post-exam analysis meeting, and details for the moderation and review measures for all assessments. The exam policies documents specified references for some statistical values for post-exam analysis, such as specifying the acceptable values for facility index, discrimination index, and reliability coefficient, which should be above 0.7 for the exam to be considered reliable. Furthermore, this document detailed some feedback issues, such as the process of receiving feedback from students after each exam, staff giving feedback to the cohort within six days after the exam, and reporting the recommended actions that need to be considered for the previous and upcoming exams.
3 & 4	OSCE guide and OSCE pro forma	3 in total	Brief checklists for the OSCE designers to follow during the preparation and implementation of the OSCE. It includes: • Creating and testing the blueprint. • Designing marking scheme. • Logistic preparation. • Conducting pre- and post-exam analysis.	This school has two worksheets, one named OSCE guide and the other named OSCE pro forma. Both documents are short and concern the same purpose of checking the steps of designing the OSCE, serving as a checklist for the OSCE designers. They both mention similar things like ensuring proper blueprinting, the need for pre- and post-exam analysis, giving feedback to students, receiving feedback from students, and ensuring logistics availability. However, having two checklists for the same purpose could be potentially confusing.

			• Receiving feedback from the students. • Providing feedback for the cohort.	Although they overlap in most points, neither is comprehensive. Merging the two into one with sufficient details and a carefully designed document would probably be more effective.
5	OSCE blueprint	1	An example of a blueprint table for one clinical course. It is a matrix table that contains the topics needed to be assessed against the psychomotor skill needed to be performed.	It is a simple table that divides psychomotor skills into three domains: history-taking, physical examination, and procedural skills. However, it does not mention other domains like communication skills, data interpretation, or diagnosis and management skills.
6	Template for a full OSCE station (Appendix 22 contains a part of this document)	7	An example of a complete OSCE design for a given OSCE station. It includes: • Template for inventory. • Examiner guide and instructions. • SPs instructions. • Students' sheet that contains the station scenario and assignments. • Rubric guide for the marking scheme. • Marking scheme. • Pre-exam review comments. • Post-exam review comments.	The OSCE designers prepare one document for each station in the OSCE exams. This document is sent to the OSCE Committee for pre-exam analysis. Then, the OSCE Committee produces a report detailing all suggestions and comments that need to be considered before approving the OSCE.
7	A station scenario and student questions	1	An example of a student sheet containing the scenario and tasks.	The scenario was clearly written, followed by four assignments which some might argue is too much for a ten minutes station (Appendix 22, second page).
8 & 9	Marking scheme	2 in total	A four-point Likert scale scoring sheet is divided into three domains with a total score and a global rating.	It appears to be a comprehensive and detailed scheme.
10	Template for inventory	1	A simple template needs to be filled out for each OSCE station to allocate the needed equipment. It includes the station's: • Domain. • Title. • Target group.	It is a straightforward inventory template.

			• Date. • Duration. • SP role. • Equipment required.	
11	Template for SPs instructions	2	Written instructions for SPs detailing how they should present the medical condition and answer students' questions.	It is well-detailed; however, it does not indicate how much training the SPs need, when it will be provided, and by whom.
12 & 13	Pre-exam analysis review (One page is appended in Appendix 23)	5 in total	A report from the head of the OSCE Committee contains recommendations and comments regarding pre-exam analysis for some OSCE exams. It includes: • Ask the OSCE designer to modify some stations. • Urge the OSCE designers to include more instructions for the examiners to ensure greater consistency in scoring. • Using a four-point Likert scale instead of a three-point Likert scale in the marking scheme. • Ask the OSCE designer to add a global rating beside the total score. • Encourage the OSCE designers to send the instructions to students and examiners sometime before the exam. • Invite the OSCE designer to utilise the OSCE guide that the school produced. • Instruct the OSEC designer to increase station time to 10 minutes.	This medical school emphasises the OSCE pre-exam analysis to check on stations' plans and preparations. The documents indicated that they run these review sessions for every OSCE exam and ensure that all stations comply with the school's standards and regulations. In the pre-exam analysis, the staff collect and analyse OSCEs' forms and templates, such as the blueprint form, the template for designing the stations, the template for the equipment needed during the stations, the template for the SPs' instructions, and the station's scenarios for the students. I collected and analysed some reports for the pre-exam analysis, found them well-detailed, and provided direct suggestions and comments for every OSCE station. Some comments, for instance, asked to detail examiners' instructions for specific stations, modify the marking scheme to include global rating, and use the OSCE template document for each station to complete all station requirements.
13 & 14	Students feedback forms	2 in total	Simple forms were prepared for students to provide feedback to the OSCE designers on the OSCE that they encountered.	The name and ID numbers are optional so that students can remain anonymous.

16	Outside examiners' feedback forms (One example is appended in Appendix 23)	3	A simple form was prepared for the outside OSCE examiners to provide feedback on the school's OSCE.	Each feedback sheet has two to three points for the OSCE implementation and future recommendations. Each station's feedback should be filled out on a separate form. The outside examiners' comments were generally very positive.
17	Post-exam analysis review	4	A report was written by the OSCE designer and the head of the OSCE Committee after the OSCE. It includes a: • Summary of the feedback received. • Brief statistical review. • Recommendations.	For the post-exam analysis, the OSCE designer and the OSCE Committee write a review of the exam based on the students' results, feedback from the students, and feedback from the internal and outside examiners. The report includes a brief statistical review that includes the students' pass percentage, the inter-rater reliability, and the mean score for each station. The committee may recommend some actions for some stations, such as removing a station if considered necessary based on panel consensus and providing recommendations for future exams.
18	A response for external evaluator	3	A part of a response written by the head of the OSCE Committee in response to an external evaluator.	This report includes a suggestion from an external reviewer to improve their reliability and validity. The head of the OSCE Committee wrote a response to this suggestion, including an action plan that included the following items: the need for a blueprint for all OSCEs, the introduction of a new standard setting, the expansion of the number of stations, and the conduct of psychometric analysis.
19	Station bank	1	A list of the stations available in the school's station bank. It includes 26 station titles.	It appears to be a small station bank. However, the school just started constructing it while still using the international medical school partner station bank.
20	OSCE workshop material	3	An outline for the workshop content prepared for the faculty development session about the OSCE.	The material provides a broad overview of the OSCE but is not detailed. It can be considered a starting point to orient the staff on designing and preparing for an OSCE exam.

5.2.1.1 The codebook analysis matrix

Using codebook analysis, I compiled a set of codes derived from the three international organisations' assessment documents and guides. Then, I created a comprehensive matrix comparing all codes found in the resources of the three guides to those found in this medical school's documents. Following this approach, four key areas emerged: 1) assessment policy, 2) criteria for good assessment, 3) use a system of assessment, and 4) educational resources for assessment—each theme contains multiple codes, fifty-eight in all. Table 15 presents the codebook analysis matrix. Most of the codes are present in all documents; however, as shown in Table 15, a few are missing from this school in comparison to the literature codes. Appendix 15 contains a worked example from the documentary material that depicts the code generation process. In addition, the table in Appendix 16 demonstrates how the codes were refined and then combined into a single area.

Table 15. This matrix illustrates the codebook analysis results for case study B's assessment documents in relation to the three organisations' assessment documents.
The underlined points indicate a discrepancy between this school's documents and the literature.

Key areas	Codes from literature Ottawa conference consensus + ASPIRE criteria + WFME standards	Codes from this medical school The assessment documents of the case study B	General guidance for the codes
Assessment policy	• Assessment policy should be responsive to the context	• Exam policies–two documents • Founded assessment unit • Founded OSCE Committee	The school assessment policy should govern and tailor the assessment according to its context and available resources.
	• Specified educational outcomes	• Course specifications for each course	The assessment plan should indicate specific educational outcomes expected from the assessment process.

	• Designing & conducting the OSCE
	  ○ Marking scheme
	  ○ Writing the stations
		• Designing & conducting the OSCE
		  ○ Marking scheme
		  ○ Templates and forms to write the instructions and scenarios
		  ○ <u>Unobserved OSCEs, which are stations with no assessors (only written questions)</u>
			The elements and steps to design, implement, and review the OSCE.

	• Appeal policy	• Appeal policy	Detailed account for petition and challenge for the assessment results.
	• Remediation process	<u>A lack of evidence in the documents sampled.</u>	Detailed account for the re-sitting of the exams for students.

Criteria for good assessment	• Validity
	  ○ Content blueprinting
	  ○ Quality assurance
		• Validity
		  ○ Content blueprinting
		  ○ Pre- and post-exam analysis (quality assurance)
		  ○ Content experts' revision
			The assessment measures what it is supposed to measure and reflects the curriculum.

	• Reliability
	  ○ Number of the stations
	  ○ Length of the stations
	  ○ One assessor is recommended
		• Reliability
		  ○ Number of the stations
		  ○ Length of the stations
		  ○ Psychometric analyses
			When the same assessment is introduced again in similar circumstances, it should produce similar results.
			
			Multiple issues mentioned in the documents contribute to the reliability of the OSCE, such as

	○ Standard-setting (criterion-referenced) ○ Correlation with other exams ○ Psychometric analyses		sampling for the stations and calculating some reliability values like inter-rater reliability.
	• Fairness ○ Test security ○ Examiner training	• Fairness ○ Controlling design & application ○ Oversight the exam ○ Examiner training	The OSCE should use a fair grading system and be free of deliberate bias, cheating, or any other issue that advantages some students over others.
	• Educational impact ○ Feedback to the students ○ Formative exams	• Educational impact ○ Feedback to the students	The educational benefit that students gain from taking the OSCE. For example, providing feedback for the students and providing formative OSCEs.
	• Catalytic effect ○ Feedback to the school and teaching system ○ Continuous improvement	• Catalytic effect ○ Feedback to the school and teaching system ○ Continuous improvement	The motivation that all stakeholders gain from conducting the OSCE. For example, the feedback that the school received to support and enhance future education, review the implementation process for quality assurance purposes and improve future implementation.
	• Feasibility	• Feasibility ○ Station bank	The assessment should be practical in terms of time, funds, and workload in a given context.
	• Acceptability	A lack of evidence in the documents sampled.	Different stakeholders—faculty and students—recognise assessment importance and credibility regarding its content and process.
	• Equivalence	• Equivalence	The assessment results lead to similar or equivalent decisions across equivalent cohorts or institutions.
Use a system of assessment	• Comprehensive ○ Use multiple tools for assessments	• Comprehensive ○ The documents mentioned only one tool for clinical assessment	The OSCE should not be used alone as a clinical assessment tool, so various assessment methods need to be employed.
	• Coherent	A lack of evidence in the documents sampled.	It emphasises that different assessment tools in the assessment system should be consistent with each other and aligned around the same purpose.
	• Continuous	• Continuous	The assessment system needs to be continuously improved and evaluated against quality criteria.
	• Purpose-driven ○ Formative and summative ○ Assessment role	• Purpose-driven ○ Formative and summative	It indicates that the assessment system should include formative and summative exams to improve

			students to graduate competent doctors who can contribute to society's health and patient safety.

	• Transparent
	  ○ Utility indicators
	  ○ Assessment supported by the assessment literature
		• Transparent	It required the school to make its assessment procedures accessible to all stakeholders to endorse accountability and integrity principles. Transparency should encourage the school to use evidence from the literature to make good assessment decisions.

	• Acceptable and appropriate	• Acceptable and appropriate
		  ○ Feedback from students
		  ○ Feedback from examiners
			The assessment system should be designed to be acceptable and appropriate for all stakeholders.

	• Feasible	• Feasible	The assessment system needs to be efficient and practical in terms of the fund, time, and workload regarding the school context.

	• Quality control
	  ○ Feedback to school
	  ○ Evaluation and analysis of performance
	  ○ Continuous improvement
	  ○ Scholarship promotion
		• Quality assurance
		  ○ Feedback to school
		  ○ Evaluation and review for each exam application
		  ○ Statistical analysis
			The school needs to ensure its assessment quality by evaluating current applications to improve future applications. This can be done by evaluating the assessment utility, which is a combination of the items mentioned in the criteria for good assessment.

Educational resources for assessment	• Human resources
	  ○ Experts for design and support
	  ○ Examiners
	  ○ Patients and SPs
		• Human resources
		  ○ Assessment experts
		  ○ Outside examiners
			The assessment documents emphasise the need for assessment experts, examiners, patients, and SPs.

	• Stakeholder needs
	  ○ Faculty development
	  ○ Examinees' needs
	  ○ Healthcare system and regulators' needs
	  ○ Patients' needs
		• Provide support for the staff
		  ○ Faculty development
			All stakeholders should receive enough familiarisation and training for the OSCE as a tool for assessment.

	• Physical resources
	  ○ Venue
	  ○ Models
	• Fund
		• Physical resources
		  ○ Venue
		  ○ Required tools
			The requirements to meet the physical and financial requirements of the assessment.

5.2.2 Analysis of the documents' findings

This section compiles and analyses the key findings of the document analyses presented in Table 14 and Table 15. It is structured in accordance with the broad subheadings of the validity framework workbook for the OSCE described by Boursicot et al. (2019, 2022).

5.2.2.1 General comments on the assessment documents

Table 15 shows that almost all the essential elements of good assessment criteria published in the relevant literature are mentioned in these school assessment documents. The available data findings suggest that their OSCE and assessment system are largely aligned with the criteria for good assessment. However, some issues have not been explicitly mentioned in the school documents, such as the acceptability of the exams, the system of assessment's coherence and feasibility. Aside from documents, other data sources may need to be considered to determine whether these issues are conscious in this school. Nonetheless, these documents were designed in such a way that it appears the school is applying these principles to some extent. For instance, they have multiple extended tables that overview the entire assessment system used in the school, implying that they consider the coherence principle.

5.2.2.2 Intended use of the OSCE in the documents

In no document are the OSCE's purpose and intended use stated explicitly, nor is it indicated how the OSCE fits into their assessment system. The documents' contents, however, imply that the OSCE is used as a summative clinical assessment and that each course has its own OSCE. However, documents show that the OSCE is the only clinical assessment tool used at this medical school, with no overarching assessment strategy that indicates the need for other clinical assessment instruments. Furthermore, the documents do not specify which competencies the OSCE should assess, only that the school should adhere to the SaudiMEDs list of clinical competencies (Alrehaily et al., 2022).

5.2.2.3 Relevant data to support the intended use of the OSCE

Planning of content

The exam policy emphasises that OSCE designers must follow these general guidelines to ensure acceptable levels of validity, reliability, educational value, and overall exam quality. However, the documents do not provide a comprehensive description of how the OSCE should be structured in accordance with evidence-based practises. While there is one example of sampling and blueprinting for an OSCE, most of the OSCE documents available are checklists and forms for each step of designing and implementing the exam to assist OSCE designers. The OSCE Committee offers suggestions for improvement in

the pre-exam analysis and generates a report after each analysis session to resolve any flaws and ensure high assessment quality.

Internal structure

This is related to reliability; one pre-exam analysis report advises the OSCE designer to increase the number of stations from five to six or seven. However, no policy specifies how the number of stations should be determined in order to produce a reliable exam, aside from the requirement that Cronbach's alpha be at least 0.7, which has not been calculated for any OSCE exam. However, it specifies that station durations be ten minutes or longer if necessary. This medical school's marking scheme example indicated that they use a 4-point Likert scale with a total score and global mark. They provided me with a station scenario that appeared concise and clear, followed by four tasks for the students. However, this could be seen as excessive depending on the nature of the specific tasks for a ten-minute station (Appendix 22). This demonstrates the absence of a system for piloting stations to ensure task-time agreement.

Response process

For the OSCE conduct process and to ensure fairness, these documents urge OSCE designers to provide detailed instructions for examiners, SPs, and students. According to the documents, the OSCE staff are trained in the design and administration of the OSCE, as well as in how examiners should behave during stations. The training aims to familiarise examiners with the exam's format and mark sheets in an attempt to create a shared understanding among them. However, I was unable to locate detailed information regarding the training procedures, duration, instructions, or whether the training session is mandatory for OSCE designers, examiners, and SPs.

Consequences of the OSCE

The cut-score is an important issue that is mentioned only briefly in the documents. The only standard-setting mentioned is that passing the exam necessitates a total score of sixty per cent of the OSCE. This is regarded as a fixed score on all exams taken at this school, with no mention of alternative, criterion-based, standard methods that can be used. This appears to contradict best practises guidelines in the literature for OSCE standard-setting (Homer et al., 2016). The school asserts that its moderation policy ensures that assessment criteria are consistently applied and that the assessment outcome (e.g., grades) is fair, valid, and reliable. The available documents do not, however, describe how the school manages the remediation exam when a student fails the OSCE.

Another important outcome of the OSCE is feedback to the student, which is asserted in multiple documents to improve the educational value of OSCEs. However, the feedback given to students is not specific or personalised but rather general for the cohort.

Effective feedback to students ought to be personalised, specific, detailed, and actionable (Boud, 2015; Hattie and Clarke, 2018; Dawson et al., 2019).

A post-exam analysis is performed after each exam to ensure the quality of assessment practices. According to the document analysis, this faculty is keen to receive written feedback from internal and outside examiners as well as students. The documents indicate that post-exam analysis would take inter-rater reliability, internal consistency, and difficulty index into account to statistically ensure the quality of the exam item. The OSCE Committee may decide to take action, such as removing a station and increasing teaching hours, based on the post-exam analysis. However, no documents show a proper psychometric analysis.

5.2.3 Summary

The OSCE documents of this medical school align with the majority of the good assessment criteria outlined in the assessment literature. Evidence suggests that this school aims to produce OSCEs that are valid, reliable, fair, and educationally beneficial by conducting analysis sessions and utilising feedback from various stakeholders. However, some practises go against the literature's recommendations, such as unobserved OSCEs, fixed standard setting, and a lack of individual feedback to students. As a result, there are several omissions and a lack of evidence-based practise in some documents, suggesting that their assessment documents require further development.

5.3 Interviews and focus groups findings and analysis

Using RTA to analyse this case study's interviews and FGs allowed me to generate four main themes (which are shared by both case studies, as explained in Section 3.8.3):

1) Institutional and assessment culture: The influence of institutional and assessment culture on the OSCE implementation.
2) Faculty expertise and practices: The effect of faculty expertise and practice on the current OSCE implementation.
3) OSCE quality and design: The current OSCE procedures impact on the OSCE quality and design.
4) Resources and infrastructure setup: The impact of resources availability on the OSCE implementation.

Table 16 summarises the themes and subthemes derived from case study B's interviews and FGs. Each theme consists of multiple subthemes, a total of ten, some of which have subheadings, summarised in Appendix 35. Due to the integrative nature of the qualitative case study analysis, there may be limited overlap between the themes and subthemes. Hence, each theme or subtheme must be considered in relation to others, making analysis of them in isolation impossible. Therefore, at the conclusion of this chapter, an overarching theme will cluster the concepts found and analysed in this case study.

Table 16. A summary of the themes and subthemes derived from the interviews and FGs.

#	Themes	Subthemes
1	Institutional and assessment culture	The nature of communication and collaboration
		The Influence of assessment regulations
		Desire for improvement
2	Faculty expertise and practices	Faculty awareness of the OSCE
		Faculty efforts in planning and preparation
3	OSCE quality and design	Attention to quality
		Theory-practice gap
		Reliability issues
4	Resources and infrastructure setup	Human resources
		Educational and logistical resources

5.3.1 First theme—Institutional and assessment culture

This theme is exploring the influence of institutional and assessment culture on the OSCE implementation. It is widely agreed that the environment of the institutional culture and social interaction would greatly influence the assessment implementation (Holzweiss et al., 2016; Simper et al., 2022). Therefore, this theme explores the factors and issues that shape the culture of this institution's assessment and influence on the OSCE implementation. I divided the aspects that affect how the OSCE is being utilised in this school into three subthemes (Table 17).

Table 17. The subthemes of the first theme and their description.

Subthemes	Description
The nature of communication and collaboration	Explores the aspects of internal and external communication and collaboration among various stakeholders that affect OSCE implementation
The Influence of assessment regulations	Examines the influence of external and internal regulations on the OSCE's implementation
The desire for improvement	Demonstrates the faculty's flexibility and desire to improve their OSCE in the future

5.3.1.1 The nature of communication and collaboration

This subtheme explores the nature of this school's attitude towards collaboration and communication with different parties. It discusses this issue within the school, the nearby medical schools, an international medical school, and the nearby hospitals and how that impacted OSCE implementation.

Internal communication

This section illustrates the nature of the communication and interaction between different units, faculty to staff, and among staff members within this medical school and its impact on the OSCE implementation. The academic leader shows that there are different units that participated in assessment implementation, and he thinks that the units work collaboratively.

> *'In Health Education Unit, we work with the Assessment Unit to consider and implement the assessment methods. We also work together to update regulations and policies to keep up with the new advances in assessment.' (Interview1-AL).*

The HPE in this medical school is equivalent to the Medical Education Departments in other schools. So, the HPE supervise and regulate educational activities—classically including assessment—in this school. This school, however, has a separate assessment

unit responsible for reviewing and approving assessment standards. Also, they have another unit for the OSCE—the OSCE committee—as discussed in the next subtheme. While this could be seen as collaborative work from different units to implement the OSCE, this could cause overlap between different units' responsibilities.

The collaborative view is supported by an assessment specialist who agrees that the communication between different parties working on the OSCE is efficient and highlights the importance of this in implementing good OSCEs.

> *'So that link is there. The administration, the PDQA (Professional Development and Quality Assurance), and the Health Education Unit collaborate with the OSCE Committee. We are all working together to continuously look into the process of, you know, making sure that the OSCE assessments are meeting high standards.' (Interview2-AS).*

Another participant indicated that whilst implementing the OSCE is challenging, the units have a good level of communication and collaboration to overcome the OSCE demands.

> *'This is teamwork, actually; no one of us can do it all. We collaborate, from writing the scenario all the way to arranging all material for the stations, conducting the exam, and lastly, compiling the results. So, I feel that everybody has taken part in these steps.' (FG4-E1).*

Additional communication channel highlighted by the academic leader was with the students.

> *'The students have a direct channel with the chief academic supervisor through the student council and the students' support club. They can request anything. Then, we can see what we can do for them.' (Interview1-AL).*

This suggests that communication between stakeholders is present, which can aid in hearing others' perspectives and resolving issues. The existence of efficient communication channels would aid in resolving problems and facilitate a more effective OSCE implementation (Banihashemi, 2011).

External collaboration

This medical school engages in external interaction and cooperation with other medical schools and hospitals for the purposes of the OSCE. This can affect how this school's OSCE is implemented, for example, by obtaining outside examiners.

> *'We have internal and external examiners for our OSCE. So sometimes we bring examiners from (medical school x) or from the hospital… This helps evaluate our students' performance and also our examiners' performance.' (Interview1-AL).*

Outside examiners' assistance benefits this school's OSCE by covering the shortage of examiners and providing feedback for the OSCE organisers.

> *'Due to the shortage of examiners, especially for some specialities like gynaecology, we invite examiners from (medical school x) or clinicians from different regional hospitals, which happened a lot. They are usually very*

supportive, and we ask them for feedback about the stations' organisation, scenarios, and so on.' (Interview2-AS).

Student exposure to outside examiners would help in preparing them for future practice in hospitals. However, outside examiners could be unaware of the students' level or the OSCE exam process. For instance, clinicians from hospitals may not be mindful of how students have been trained to perform skills and techniques. Therefore, students might be assessed according to these clinicians' expectations. Subsequently, some outside examiners may assess students differently, which would affect the reliability and validity of the OSCEs, especially in parallel OSCEs. On the other hand, outside examiners may also provide feedback on this medical school's OSCE practice, which would incorporate outsiders' opinions to develop the exam.

The administration's support is critical for a successful assessment implementation (Weiner, 2009). The administration of this school facilitated external collaboration, suggesting that the leaders of this school are aware of their role in facilitating OSCE implementation.

'I think the administration is very supportive and facilitate things as much as they can, for example, in inviting outside examiner from some specialities.' (FG3-E2).

The faculty is considering a project that has the potential to improve cooperation across universities in the area. Because they have an availability issue with the simulators and SPs, they want to overcome this challenge by collaborating with other institutions in the area.

'We keep thinking about the possibilities of having a pool of trained simulators available for our OSCEs, which we can make with other institutions as assessment logistics.' (Interview3-AS).

This medical school also collaborates with their international western partner to acquire OSCE scenarios.

'We receive some of the OSCE stations and the assessment forms from (the partner medical school), but sometimes they do not fit our cultural or the local context. Therefore, we are (the examiners) helping in adapting the scenarios to the local context.' (FG4-E4).

The above suggestion may make the OSCE exam more feasible, yet the participant identified that they need to re-write some stations for their authenticity. I remember one example this participant mentioned during a side talk after the FG about a station that they received from their western partner. The station was about dealing with a drunk man brought to the hospital with a medical condition which is a very rare case in SA culture, so the local OSCE writers need to re-write the scenario to make it more authentic and culture friendly.

In summary, according to data, this medical school's culture is characterised by multiple levels of communication and collaboration with various stakeholders, suggesting that all stockholders can communicate their voices and support each other. The faculty are aware of the importance of teamwork in providing suitable OSCEs, as efforts can synergise to make successful and feasible OSCEs (Thornhill et al., 1996; Suter et al., 2009). Yet the analysis indicates that multi-units controlling the OSCE implementation might add unneeded complexity and hinder faculty communication which may lead to challenges in OSCE implementation.

Externally, this medical school shows examples of communication and collaboration with medical schools, local hospitals, and international institution, which benefited their OSCE implementation (Heal et al., 2019). They invite outside examiners and receive validated OSCE stations from recognised third parties to implement feasible and valid OSCEs. However, the faculty recognises the need for caution when incorporating external OSCE stations.

5.3.1.2 The Influence of assessment regulations

From a cultural point of view, rules and regulations are potent symbols which usually form, represent, and direct the culture of institutions (Ndoye and Parker, 2010; Simper et al., 2022). This subtheme reveals the impact of these regulations on the OSCE implementation in this medical school. I categorise the regulations into two levels, external regulation to the level of regulatory bodies and internal regulations to the level of this medical school.

External regulations

The interviewees' response to a question about how they make choices during the different implementation stages and what is the regulatory bodies' influence on that. The following quotes show that they adhere to external regulators.

> *'When we are preparing, as we said, the blueprint of the skills for the exam… they are built according to the national standard of the SaudiMEDs framework, so the blueprint is directly influenced.' (FG1-D1).*

> *'… the NCAAA asked us about the reliability of our OSCE.' (Interview1-AL).*

The NCAAA accredited this medical school. So this medical school needs to prepare an annual progress report to follow up on any significant changes to their programme and submit it to the NCAAA (ETEC, 2023b). The NCAAA uses SaudiMEDs as a framework for the competencies that graduates should achieve (1.3.2). This shows the direct impact of the NCAAA and the SaudiMEDs on the OSCE in this medical school. This could be used as validity evidence that this faculty follows a well-established competency framework during the OSCE implementation. Although this medical school can overlook these organisational bodies' instruction, the medical school will lose its accreditation

which will affect its reputation. This would affect their graduates' employability and the number of new students joining this medical school. Furthermore, some authors argue that obtaining NCAAA recognition would enhance the quality of medical school education and facilitate international recognition for their graduates (Alrebish, 2014; ETEC, 2022). Therefore, most medical schools in the KSA are trying to comply with NCAAA regulations to gain accreditation and the beneficial consequences of that recognition.

> *'We responded very well to the NCAAA recommendations. For example, the last recommendations were to focus on students' performance on each point in the OSCE rubric to detect the issues in the curriculum and teaching. They also emphasise the importance of the post-exam analysis. We accept and implement these comments and find them useful. This will help us to improve next year's application and improve our teaching for clinical skills.' (Interview2-AS).*

This school seems responsive and flexible in accepting and incorporating new ways of dealing with their assessment to improve its implementation. The faculty saw that as an opportunity to enhance their practice. Therefore, the accreditation agencies' recommendations are influencing the practice in this medical school because they need to follow them to obtain accreditation. This could be seen as a pragmatic way to gain the benefits; however, all medical schools have to do so to get accreditation from this agency.

Internal regulations

Interviews revealed that participants were satisfied with the clarity and appropriateness of their OSCE assessment policies. One assessment specialist commented:

> *'I think the college regulations for assessments as a whole and particularly for OSCEs encouraged us to maintain high standards. It gives us a lot of flexibility and encourages us to improve. The administration always asks us to maintain the quality cycle… and reflect on our assessment, learning, and teaching.' (Interview3-AS).*

The academic leader was pleased with their policies and their responsiveness to change according to their circumstances.

> *'Our policies are all clear and updated. So, I believe we regularly change our rules and regulations based on our needs. And if we find something negative, we put it in our improvement plan. Sometimes the issues that we need to change are based on something that we feel needs to be updated, from the students, the faculty, and the Saudi regulatory parties like the NCAAA or the ministry of education… The head of the Assessment Unit is the one who works on and is entirely responsible for the assessment guidelines.' (Interview1-AL).*

The assessment unit includes one person responsible for reviewing and approving assessment guidelines. However, forming the assessment guidelines for the whole medical school by one person would be undesirable. Because if this individual is misinformed, that will have an impact on the whole assessment system. Furthermore,

from the document analysis, it was not clear how they constructed their guidelines, as no references have been mentioned in any of the documents received. Another possibility of having only one person in this unit could be due to financial restrictions. Since this is a private medical school with limited budgets, staffing may be affected, as discussed in the fourth theme. However, for the OSCE, the faculty formed an OSCE Committee dedicated to overseeing and ensuring standardisation in the OSCE implementation across all courses by adhering to specific and unified criteria.

> *'OSCE Committee's responsibility is to build up one strategic plan for the whole clinical rotation to standardise the OSCE. So, internal medicine would not be very different from surgery and so on. They have sent us specific forms to be used during the preparation of the stations, different files for students' instructions, examiners' instructions, checklists, rubrics, and so on.' (FG1-D2).*

The participant sees the OSCE Committee as a central quality assurance measure. However, the analysis of the documents at the beginning of this chapter does not show detailed OSCE-related regulations. The general assessment regulations presented in the documents do not consider the uniqueness of the OSCE nor provide a comprehensive OSCE regulation. Thus, the OSCE forms given by the OSCE Committee are not solely enough to implement good OSCEs. Therefore, staff experience is particularly crucial for the successful implementation of OSCEs. One OSCE designer said:

> *'Because we have been doing the OSCE for some years now. And there are two to three courses per year. So, due to this experience, the OSCE organisation is much better now.' (FG2-D3).*

This implies that the guidelines are constantly refined based on experience and practise. However, the following example suggests the clarity of general assessment policies and the staff's familiarity with these regulations. It shows a clear policy for applying the fixed cut-score rule for different study levels.

> *'The pass cut-off score is sixty percent; it is fixed. For the first three years OSCE, the Clinical Skills course is a standalone course, so students must get at least sixty per cent on the whole course to pass. But for the last three years of the school, the OSCE is part of the assessment programme for each clinical course; the OSCE score is aggregated with other exams scores to reach sixty per cent or above to pass each clinical course.' (Interview2-AS).*

Although this assessment specialist appears to know some fine details about the cut-score rules in this school, a fixed pass score goes against best practices for high-stake OSCEs (Pell and Roberts, 2006; Park et al., 2021).

In summary, the data indicate that this medical school abides by the regulations of the regulatory bodies, suggesting that they adhere to high standards and regulations in their OSCEs. Efforts are being made to standardise the OSCE examination experience across the institution, which will enhance the OSCE's validity and produce defensible results.

However, excessive standardisation and control would not reflect real-life practice and could make it difficult to meet the specific OSCE needs of different years and courses. Therefore, a level of flexibility might be needed to embrace assessment creativity and innovation. Furthermore, the findings from the documents analysis indicate the absence of comprehensive OSCE regulations. The analysis suggests that the faculty working on the OSCE are well informed by the assessment regulations' details, which propose good communication within the faculty. Yet, some examples indicate that this school regulation might not be supported by evidence-based practice; therefore, the claims of having updated regulations might not be accurate.

5.3.1.3 The desire for improvement

This subtheme demonstrates the tendency for improvement that can be noticed among the faculty. This impression is supported by several quotes suggesting its prevalence in this medical school's culture. This subtheme is characterised by two aspects: flexibility and adaptability and staff vision for development.

Flexibility and adaptability

This highlights the faculty's attitude toward change and development.

> *'In (medical school x), we believe everything needs to be updated and improved. We are receptive to learning and improving things. This is life. No one is perfect.' (Interview1-AL).*

The excerpt above suggests that some staff are open and adaptable to change by consistently incorporating feedback into their practise. However, the important issue is on what basis this change will be introduced. Change must bring about improvements based on evidence-based evidence rather than change for the sake of improvement. The third theme—under the unobserved OSCEs section—gives an example of how the faculty introduce change.

The faculty are receptive and appreciative of external regulatory bodies' feedback.

> *'We are using the feedback that we had from the NCAAA review visit last year. We, you know, took those recommendations very seriously. We are thankful to them for giving us this feedback.' (Interview3-AS).*

Also, they collect feedback from four different sources from their medical school, as this assessment specialist described:

> *'We put in our consideration the feedback that we received from the examiners, from the coordinators, from the committee of the OSCE exam itself… and from the students also.' (FG1-D5).*

There is an incremental improvement to the OSCE, which shows the ability of the faculty to introduce change.

> *'Every year, we introduce one or two improvements to the OSCE stations. So again, it is an ongoing cycle for us; based on the feedback we receive to*

improve things… for example, we received good feedback on the checklist, then we have modified it.' (FG4-E5).

The staff are ambitious and looking for further improvement, often by learning from the practice of other medical schools.

'If you go through some prestigious universities, they have printed guidelines for OSCE exams for students. Everything is very clear, so I think we need something like this. Things need to be clearer. However, we are gradually overcoming some problems. We had a problem with the rubric, and we had a problem with the checklist also. But all those things have been gradually solved.' (FG4-E1).

This participant identified the need for solving such issues by, for example, learning from well-known institutions' experiences. This also highlights their adaptability in resolving issues with their marking schemes.

The data suggest that this medical school is receptive to feedback and continuously incorporates that to improve the OSCE application. They are open to hearing from different stakeholders, so they are involved in the development of the OSCE, which would increase the acceptability of the OSCE among them. The quotations suggest that this medical school can readily embrace and implement change and keep improving toward higher quality OSCE. This may be affected by the fact that the medical school is relatively young; therefore, it is not bound by history or outdated, static systems. This helps shape a positive and expansive culture in this institution, allowing for innovation, adaptability, and continuous quality improvement (Simper et al., 2022).

Vision for development

The participants discussed some areas for potential improvement of the OSCE. An OSCE designer demonstrated his enthusiasm and eagerness to implement a high-quality OSCE by working to enhance its reliability and validity.

'If we can achieve the reliability, validity, and objectivity of this exam, then the challenges we face should not matter… But if we give up and introduce low-quality OSCE that cannot meet the main goal of this exam, I think this will be the main problem, and the exam will be a waste of time... Therefore, we are trying to achieve the best OSCE either by considering the feedback, by reading in the literature, and by learning from our experience.' (FG1-D5).

The participant inferred that the current implementation is not optimal, but he demonstrated commitment to a better implementation. Furthermore, one assessment specialist aims higher, thinks about large-scale development, and collaborates with other institutions.

'We also need to think bigger by merging the Assessment Unit with the OSCE Committee to establish an Assessment and Examination Centre that helps not only our college and university but also other universities in the region… so we can get more assessment experts and OSCE examiners as well as making profits.' (Interview2-AS).

While this suggestion is supported by some literature (Allan et al., 2023), the faculty shortage and financial constraints, as described in the fourth theme, may render the formation of an assessment centre unattainable. However, since this is a private school and this initiative has the potential to generate revenue in addition to its educational value, they may find a means to implement it. Another participant identified specific areas that he desired to improve next.

> *'We are looking at the possibility of doing a proper psychometric analysis, including the individual items in the checklist. That is what we want to achieve next, Insha'Allah.' (Interview3-AS).*

One assessment specialist wished they could conduct a 'proper' psychometric analysis. And conclude the dialogue with '*Insha'Allah',* which literally means 'If God wills' to affirm his desire. He indicated later that was one of the non-compulsory suggestions from the NCAAA. Although that was an optional suggestion, he seemed truly want that which may show his fidelity to his workplace. A few simple statistical parameters are already calculated for the OSCE—discussed in the third theme. However, he was aiming for a more sophisticated analysis, yet it appears that they have no psychometricians. Another suggestion was to consider an alternative standard setting.

> 'We are using the pass marks as a sixty per cent, fixed one… Do we have a system here? But I think that still needs to be changed.' (FG1-D3).

However, the staff generally appeared content with their OSCE yet keen to enhance their practice.

> *'We are quite comfortable and happy with the (OSCE) standards that we are trying to improve, but we always like to work for, I believe the word is (perfection). So, we need to keep trying to do better and better.' (Interview3-AS).*

The findings suggest that the participants tend to feel comfortable regarding their OSCEs, but they are aware that there is room for improvement. Some staff have the vision to develop certain areas of their OSCE implementation, which may suggest their engagement and commitment to the workplace (Alsharif, 2011). Thus, if they can make informed decisions and continue to improve, the OSCE quality will improve.

To summarise the findings in the first theme, the three subthemes discussed in this theme are all around one central concept: the influence of institutional and assessment culture on the OSCE implementation. The data suggest that the communication culture inside this institution and the collaboration with other institutions are effective to some extent. However, they may need some enhancement, which could help implement valid and feasible OSCEs. This school follows the external bodies' regulations, but it has its own internal regulations, which may need some revision to improve the OSCE implementation. The analysis shows that the faculty have a desire for improvement. They are adaptable and flexible in introducing change, and the staff has a vision for

development; this may allow the faculty to introduce incremental developments, leading to better OSCE implementation. While the analysis indicates that this school's institutional and assessment culture is not optimal and that it needs further enhancement to support OSCE implementation, the faculty demonstrate a clear desire for improvement.

5.3.2 Second theme—Faculty expertise and practices

This theme explores the effect of faculty expertise and practice on the current OSCE implementation. It reflects the experience the faculty currently have and have brought with them while managing the OSCE on the ground and how it shapes current OSCE conduct and delivery. It captures the reality of current OSCE practice based on participants' understanding, perceptions, and practices. This theme contains two subthemes, as in Table 18.

Table 18. The subthemes of the second theme and their description.

Subthemes	Description
Faculty awareness of the OSCE	Explores faculty understanding of the OSCE and their perception of their OSCE
Faculty efforts in planning and preparation	Understands the processes that the people working on the OSCE follow to operate their OSCE

5.3.2.1 Faculty awareness of the OSCE

This subtheme reveals how the faculty perceives and understands the OSCE's significance as a clinical assessment instrument. It also captures the opinion and perception of the faculty on the OSCE's status in their institution and how they see its present implementation.

Faculty understanding

This section aims to understand the viewpoint of those involved in the OSCE implementation about the OSCE as a clinical assessment instrument.

> *'The OSCE is unique. As in my opinion, it allows the examiner to properly and objectively examine his students.' (FG2-D1).*

> *'It is very much needed and appreciated. It is a very good way of assessing psychomotor skills from different angles. It is the best way to assess students' (clinical) performance.' (FG2-D2).*

This fulfils an important element of the criteria of good assessments, which is the acceptability of the assessment instrument. In his later speech, the OSCE designer in the second quote stated that their OSCE assesses students' performance competencies

such as history taking, communication skills, clinical procedures, and management skills. Nevertheless, one assessment specialist provided a more holistic opinion; he noted:

> *'You are never going to have a perfect assessment. The perfect assessment tool does not exist, right?… (however) the OSCE does rate relatively high, very high for its effectiveness as an assessment tool.' (Interview3-AS).*

Another participant indicated aspects that make the OSCE a favourable tool.

> *'OSCE, in general, would allow, you know, practising on patients in a safe, controlled environment, right? We can test many types of skills in it. It does not just examine procedures but also communication skills. It is a very flexible tool, so it got a lot of variation. You can also assess a large number of students in the same standard environment.' (Interview3-AS).*

These views are thoughtful and supported by the wider OSCE literature (Abdulghani et al., 2015; Harden et al., 2015). Moreover, an OSCE examiner recognised that the OSCE designers could control the level of difficulty:

> *'And we adjust the station according to the level of the students. So, mostly it is a good exam. But of course, it needs a lot of (planning) to arrange this exam. It needs a lot of work.' (FG4-E1).*

And he also recognised the price of these advantages as the OSCE is resource intensive. The OSCE designer valued the OSCE as it can be standardised in a controlled environment which could enhance the exam's fairness.

> *'… there is standardisation, so the students have equal chances to be assessed on the same thing. So, this standardisation is a very important part.' (FG1-D3).*

The academic leader indicated that the quality of the OSCE depends on the quality of its implementation.

> *'Again, it depends on how you utilise the tool. The OSCE needs a good design to be effective. So, I think the tool's usefulness depends on how it is practised.' (Interview1-AL).*

It is not only about the type of assessment instrument used but, more importantly, about how it is implemented. Additionally, an OSCE examiner regarded the OSCE as superior to traditional clinical examinations.

> *'In traditional clinical assessment, the clinical examination comes at random; the students may face random two or three cases only, and they will not cover a wide range of skills. While the OSCE covers many skills that match the learning goals because you can cover many topics.' (FG3-E1).*

The above participant appears to appreciate the significance of blueprinting and constructive alignment, thereby reducing the OSCE's reliance on randomness. He implied that the OSCE has more fairness, validity and reliability than long- and short-case examinations. Yet, he has a well-balanced insight in line with the broader literature, as he mentioned that both assessment tools are needed:

> *'Both are needed because, from the long case, you will take a history, you will evaluate the whole case as it is in real life from A to Z., But each OSCE station targeted a single learning goal.' (FG3-E1).*

In summary, the faculty responses regarded the OSCE as a valuable assessment tool; however, they believe that its effectiveness is largely contingent on the quality of its implementation. Data shows that faculty seem capable of describing what they feel is best practise in the OSCE and broadly understand many of its advantages and limitations stated in the literature (Khan et al., 2013b; Abdulghani et al., 2015; Harden et al., 2015). They believe it is a valid, reliable, fair, flexible, and useful assessment tool that can fulfil the criteria of good assessment. So, the OSCE is a widely accepted assessment instrument among the faculty.

Faculty opinion on their OSCE

The previous section explores faculty opinions about the OSCE as a clinical assessment instrument. But this section traced the participants' views and perceptions of their current OSCEs and their influence on future student practice. The faculty's perception of the current OSCE affects whether they recognise the need for improvement or lack the motivation to do so. Participants had a sense of satisfaction with their OSCE implementation.

> *'I tend to think that we are doing a great job on the planning aspect and the implementation of the OSCEs.' (Interview3-AS).*

However, these opinions are largely based on their feelings but not necessarily supported by evidence from the practice, as discussed in several sections of the third (4.3.3.1) and fourth themes (5.3.4.2). Although I have dedicated a subtheme to the faculty's desire for improvement (5.3.1.3), these opinions may render the Faculty less committed to implementing the improvement plans. Since a sense of contentment with the current OSCE implementation may impede further development. In contrast, some participants are not entirely pleased with the current OSCE implementation.

> *'To be honest, I think we are gradually improving our clinical assessment. Now, I am partially satisfied… I can say it is above the average…' (Interview1-AL).*

> *'I think it has sort of good quality but not of the top quality. But generally, regarding the checklist, rubric, and preparation steps for OSCE, I think it is of high quality; they are not bad at all.' (FG3-E2).*

When faculty members believe that there are areas in which they can improve, it motivates them to strive for better implementation, as discussed in a previous subtheme (5.3.1.3). This would also motivate them to acquire further knowledge to fill the gaps in their practice. However, some examiners argued that their OSCE is of high quality because it has an educational impact beyond the exam, that students do not easily forget the topics covered during OSCEs.

> *'I believe that the students will not forget the cases that they experienced in the OSCE, they will go and read more about the topics of the stations, and they will focus on them more in the future.' (FG3-E3).*

Other participants support this opinion:

> *'Our students are performing excellently in clinical practise… We noticed that our students perform better than their peers in other medical schools, especially in clinical questions; we believe that mostly because of the use of the OSCE in the college from year one.' (Interview2-AS).*

> *'The hospital staff said that the (the name of this medical school) graduates do have good clinical skills. Also, some feedback is coming from our graduates that when they are working in the hospital, their basic skills are excellent.' (FG4-E3).*

These claims from Faculty and hospital doctors can be regarded as validity evidence supporting the OSCE's implementation quality and educational impact. The statement from the OSCE examiner (FG4-E3) also suggests that alumni believe the OSCE helped them achieve clinical excellence. This could be because the OSCE experience prepares them for similar exams in post-graduate training. However, the statements in the preceding quotes about the OSCE's impact on future practise may be influenced by social desirability bias, where participants only show what they like (Grimm, 2010); and a lack of third-party benchmarking that assesses practise before and after graduation (Engelkemeyer, 1998). Furthermore, clinical skill proficiency is multifactorial, making it difficult to attribute it to a single teaching or assessment method. Though, Komasawa et al. (2020) and Horita et al. (2021) found that the OSCE is well correlated with future clinical clerkship performance.

5.3.2.2 Faculty efforts in planning and preparation

Planning and preparation for the OSCE consist of multiple steps and effort from the faculty. In each step, faculty choices and actions based on their expertise and practices affect the OSCE implementation. The participants described a number of OSCE preparation steps that were characterised by early planning and blueprinting, as well as the exertions of the organisers and examiners.

Early planning and blueprinting

The OSCE is a sophisticated assessment instrument that requires extensive preparation (Bearman et al., 2020). Therefore, early planning is necessary to allow sufficient faculty time to decide and write the stations, organise the needed resources, and allocate examiners and patients/SPs (Boursicot, 2010).

> *'The OSCE Committee with the Assessment Unit and the Medical Education Unit set together to design the next year's academic calendar... They book all exams dates, including the OSCEs. Then, early in the academic year, the OSCE Committee gather and decides what ILOs should be covered by*

making an extended blueprint and sending it to the course coordinators.' (Interview2-AS).

The data showed that the planning for the OSCE in this medical school starts early. This allows enough time to prepare and design the exam and avoid any unforeseen circumstances or time pressure. The OSCE Committee plays a central role in the preparation process for the basic years OSCEs.

'The OSCE Committee design the exam, allocate the time and the venue, call for the examiners, and prepare a report to be sent to the head of the Assessment Unit, head of the Clinical Department, and head of the Medical Education Unit. So, everyone knows what the exam is about, and they all can check on the ILOs coverage and so on. All parties can provide comments to the OSEC Committee.' (Interview2-AS).

But still, an OSCE designer finds that the planning and preparation for the OSCE are resource intensive.

'It takes a long time for preparation. The standardisation and training for the examiners consume so much time and effort. Also, it is expensive. These, I think some of the disadvantages of the OSCE.' (FG1-D4).

This may imply that their resources are insufficient for such an exam, as I elaborated later (5.3.4.2). Nonetheless, they appear to have no other option as they rely solely on the OSCE for performance assessment (5.3.3.2). Therefore, this faculty may need to secure enough resources for the OSCE and utilise another performance assessment instrument to mitigate the impact on the constrained resources.

A participant mentioned that blueprinting is the first step in structuring the OSCE exam, and it is done for all OSCE exams.

'In surgery. I have about fifteen skills in my course... So the students know that these are our skills. We gave them these skills in the student guidebook. We have to choose six out of these in our blueprint. We do follow a standardised blueprint from the SaudiMED.' (FG1-D3).

The use of blueprinting is a crucial step in ensuring a high level of OSCE validity (Harden et al., 2015; Raymond and Grande, 2019). However, each clinical course in this medical school has its own OSCE, so only a few skills can be blueprinted. For example, the surgery course is a major course with fifteen skills to be blueprinted, suggesting that minor courses would have fewer skills. This indicates that the OSCE exam will cover the majority, if not all, of the course's skills. This would render the blueprinting useless since all or most of the skills would be assessed. The OSCE will then be more resource-intensive when performed for each course. Moreover, this medical school has a maximum of six OSCE stations per course, which would reduce the exam's reliability, as discussed in the next theme (5.3.3.3).

After blueprinting, one clinical teacher writes the station, which is then discussed in the department before being sent to the OSCE Committee for review.

> *'The preparation and writing of the OSCE stations are sorted by the course coordinator in collaboration with the teachers in the (course) planning group... each member of the department, for example, is assigned to write one station, okay. After that, the whole group attended the assessment unit for what is called a pre-exam analysis. After that, all scenarios should be sent to the OSCE committee for final approval.' (FG1-D4).*

This means that each station is revised twice by different people, which would enhance the quality of the stations. Because writing the stations is an important component of the OSCE, it is not only about writing the scenario but also the instructions for the examiners, students, and SPs.

> *'Besides writing the station, I need to write instructions for the student. And there are instructions for the examiner, including an explanation for each point in the checklist... There are also instructions for the role-player, including which task he must perform, when to answer the question, when not to answer, and what answer he should give if asked. I finally need to make an equipment list.' (FG3-E4).*

This suggests that the faculty working on the OSCE are mindful of the need for clear instructions for all OSCE stakeholders. In addition, according to some OSCE designers, they make an effort to make the exam as authentic as possible. Indeed, the more authentic the exam, the more valid it will be.

> *'I make a simulated scenario from real-life (scenarios) to mimic real-life as much as I can, then I write the questions.' (FG2-D1).*

The academic leader indicated that a content specialist writes the OSCE, and they have an OSCE question bank.

> *'We have expert doctors in their specialities who write the OSCE. And also, we have a station bank for the OSCE.' (Interview1-AL).*

Since their OSCEs are written by content specialists, they can achieve a high level of validity (Khan et al., 2013a). And their OSCE can be more feasible because they have a bank of available OSCE stations (Abdulghani et al., 2015). Nevertheless, before being added to the station bank, OSCE stations should be thoroughly reviewed to ensure they adhere to the evidence-based design. However, it appears that piloting stations are not performed, as no document or participant mentioned this during the lengthy description of their implementation method. Piloting stations are essential to ensure smooth running during the actual OSCE operation (Abdulghani et al., 2015; Harden et al., 2015).

Duties of the OSCE organisers and examiners

In this medical school, the OSCE designers are also known as OSCE organisers or coordinators because they will be in charge before and during the OSCE exam. The participants identify several duties that the OSCE organisers and examiners need to perform just before the exam. Their role is critical to make a successful OSCE. One OSCE designer summarised OSCE organisers' role into three main areas.

> *'Our role mainly is in three parts. The first role is we design our OSCE questions and the instructions for the stations. The second part is the organisation as we organise the logistics, which also extends to training the simulators. The third part is running the exam, as we may act as examiners.' (FG2-D3).*

There is some contradiction between the participants about who writes the stations. While most participants indicated that the OSCE designers write the stations, others mentioned that the content specialist writes the stations, as in the last quote by the academic leader (Interview1–AL). This suggests that the academic leader may not be fully acquainted with the OSCE practice, which could be understandable. This, however, suggests that the administrative staff are not supervising the implementation process, which may allow for poor practice. However, if a content expert does not write the OSCE, its validity and authenticity would be diminished, though this would depend on the level of practise (Khan et al., 2013a).

Before the exam, there is a step to prepare the examiners that is different from the examiner training discussed in the fourth theme (5.3.4.1). It is common practice for the OSCE organisers to share the station scenario, station topic information, and checklist with the examiners prior to the exam.

> *'We receive the checklist and background information (about) the station that we are responsible for from the coordinator of the OSCE one week before the exam. I read the station, and if I have any problems with it, I have to be sure of how to execute it. So, I go to the OSCE coordinator or to a content expert to ask and be sure I understand how I could observe students correctly.' (FG4-E4).*

> *'The OSCE Committee... invite clinicians and non-clinicians. I am a Microbiologist, but they send me two things before the exam. They send me the checklist and background about the station to read. Because I am a non-clinician, the material they send is very useful to prepare myself for my station and fairly conduct the exam.' (Interview2-AS).*

The participants in the above quotations spoke positively about receiving the materials in advance. This suggests that this step will allow them to arrive well-prepared for the exam (Malau-Aduli et al., 2023). However, due to a lack of content specialist examiners, the faculty may resort to inviting non-clinicians/specialists. Sending the materials ahead of time is particularly important for examiners who are not clinicians, do not specialise in the assigned stations, or are outside examiners so that they can accurately assess students. While this practice is useful for examining some skills, some clinical skills require the use of content experts to assess students' performance (Abdulghani et al., 2015). However, sending stations materials to examiners in advance of exams may increase the likelihood of leaks.

Moreover, by providing exam materials before the exam, examiners could provide feedback to OSCE designers to improve the OSCE implementation.

> *'As an examiner, I usually give my comments about the exam checklist to evaluate the clarity and quality of the checklist and whether I need to add some points to clarify things for the student. My comments as an examiner are very important to the designers and will guide them in the next OSCE exam.' (FG3-E4).*

Just prior to the exam, the OSCE designers provide a briefing to the examiners to ensure they are aware of their stations.

> *'Before the exam itself, we have a calibration meeting for the examiners. And I think that is probably one of the most important things we do to have a fair OSCE assessment.' (Interview3-AS).*

> *'The good point in our OSCE is that we have a quick meeting before the exam… When I am assessing a skill, I mostly receive a briefing about different items' weight in the checklist of my station.' (FG3-E2).*

By briefing and sending materials in advance, examiners will be able to ask questions regarding uncertain points, thereby increasing the validity of their assessments (Malau-Aduli et al., 2023). Moreover, this would make the examiners consistent in their assessment across the cohort (i.e., intra-rater reliability) as well as make the examiners more consistent among each other (i.e., inter-rater reliability) as they share similar understanding, especially in parallel stations. This step will enhance the validity, reliability, and fairness of the OSCE (Fuller et al., 2017).

Exam security and supervising the application are two of the responsibilities of the OSCE organiser.

> *'… The OSCE organiser is always present to supervise exam security and ensure no contact between students during exam time to guard exam privacy; you know, this is very important. He also monitors and organises the application of the exam and ensures that everything runs as expected and that the male and female OSCE exams started at the same time.' (Interview2-AS).*

The assessment specialist noted that close supervision during the OSCE is pivotal because this medical school typically has parallel stations, one for male and another for female students (Fuller et al., 2017). As a result, it is the responsibility of the OSCE organiser to ensure that the parallel station functions properly to ensure exam reliability and security.

To summarise the findings in the second theme, each of the two subthemes discussed here revolves around a single concept: the impact of faculty expertise and practice on the current OSCE implementation. The first subtheme data suggest that the faculty believe the OSCE is a good clinical assessment instrument. They are generally pleased with their OSCE implementation. Yet, they think they want to improve its implementation. Overall, the data show that the faculty have a good understanding of the OSCE theoretical concept and believe they are implementing a good OSCE.

The second subtheme data suggest that the faculty working on the OSCE duties are early planning for the exam, blueprinting the skills, writing authentic stations, and reviewing the scenarios twice. There is a sense of teamwork in all the aforementioned steps between the OSCE designers, organisers, and examiners, which would contribute to implementing good OSCEs. The OSCE organisers send information about the stations to the examiners and brief them before the exam. While this practice allows non-specialised examiners to participate in the OSCE, exam security may be compromised. In general, the data suggests that these collective efforts would aid in the implementation of OSCEs with high validity and reliability.

5.3.3 Third theme—The OSCE quality and design

This theme examines the impact of current OSCE procedures on OSCE quality and design. It highlights the OSCE's key practises on the ground and how they influence its current implementation. This theme contains the three subthemes listed in Table 19.

Table 19. The subthemes of the third theme and their description.

Subthemes	Description
Attention to quality	Explores how this medical school ensures the quality of their OSCE implementation and the issues that affect it
Theory-practice gap	Inspects the mismatch between OSCE literature and this medical school OSCE practises
Reliability issues	Investigates the factors that influence the reliability of their OSCE

5.3.3.1 Attention to quality

This subtheme analyses the quality measures utilised by this medical school and their influence on their OSCE implementation. Some quality assurance measures include the establishment of an OSCE committee, pre- and post-OSCE review, and incorporating stakeholder feedback. Moreover, the language used to introduce the OSCE may affect the quality of their OSCE, as discussed at the end of this subtheme.

Measures for quality control

According to the discussion that follows (5.3.3.2), the OSCE is the only performance assessment instrument employed at this school. Therefore, it tries to ensure the quality of its OSCEs by monitoring their implementation.

> *'I believe that the formulation of an OSCE Committee does indicate that the institution administration pays a lot of attention to the OSCE exams, and we want to achieve and maintain excellent OSCEs… The main responsibility of*

> *the OSCE Committee was to supervise and guide OSCEs from year one to year five and to ensure that we maintain standards.' (Interview3-AS).*

This medical school established an OSCE committee to coordinate and oversee its implementation due to the importance of the OSCE in its assessment system. This appears to be an attempt by the medical school's leaders to ensure the quality of their OSCE. It supervises the whole OSCE implementation process by providing support on how to implement the OSCE and overseeing its application by the OSCE designers. So, the process is intended to be consistent across all courses. An OSCE designer was pleased with the OSCE Committee's role in assisting them in designing and implementing the OSCE in their courses.

> *'We have an OSCE Committee, which reviews the OSCE stations we design and also gives suggestions to make it more accurate and standardised. The stations have been extensively and thoroughly revised by the chair of the OSCE Committee. He gives us opportunities to discuss and challenge.' (FG2-D2).*

Although the participants see supervising OSCE implementation as a good thing, too strict supervision and rigid rules could impede innovation and creativity in the OSCE implementation (Shand, 2020; Chan et al., 2023). Furthermore, if this is not carefully balanced, it may lead to a stagnant OSCE, and the task of the OSCE designer could become limited to checking boxes and copying and pasting. However, it is reasonable to presume that none of this occurs since no participant raised this concern.

Furthermore, the OSCE Committee conducts a review of the OSCE not only before but also after the exam:

> *'The pre- and post-OSCE analysis is always conducted to ensure the quality of all our exams and to improve next year's OSCE, either by improving the teaching or the exam implementation itself. We also receive feedback from the students after the OSCE so that they can share their opinion, and we accept and amend things if we find a problem.' (Interview2-AS).*

The post-exam analysis is conducted for each OSCE exam as part of the quality control process. This may include a qualitative and quantitative evaluation (Tavakol and Dennick, 2011b). As mentioned earlier in the document analysis (5.2), they produce a short report that includes a summary of the feedback received from the internal and outside examiners and students.

> *'The analyses (post-exam analysis) actually is not just for assessment purposes, but it forms a very important tool for tweaking our curriculum and teaching strategies.' (Interview3-AS).*

This analysis benefited the current exam and was also used to improve the curriculum, teaching, and future implementation. Therefore, the OSCE results provide teachers with information regarding student learning and the effectiveness of their teaching, giving the OSCE assessment value, educational impact, and catalytic effect (Norcini et al., 2018).

Furthermore, the OSCE Committee considers the feedback received from different stakeholders, such as regulatory bodies, examiners, and students. This can be seen as a quality assurance method that could help improve the OSCE.

> *'The OSCE Committee get all points of view, examiners and students, then the good feedback incorporated for improvement.' (FG4-E5).*

The excerpt above suggests that this medical school is receptive to feedback but possibly that they are selective about what feedback they decide to act on. People tend to focus on feedback that they agree with rather than feedback that challenges them, which may be the most useful. While it is understandable that they would focus on the *'good'* feedback, other feedback should not be ignored. The committee may need to communicate with the feedback providers to understand their point of view, or the faculty can explain their view. This would increase the value of the feedback and strengthen the faculty's credibility among the stakeholders.

In addition, hearing from the OSCE examiners in the post-exam debriefing sessions is an important part of the quality assurance cycle.

> *'After the exam, the OSCE coordinators usually invite us to sit together and discuss how the exam went, so we can provide our comments… We also have a meeting with OSCE Committee to revise students' results and the feedback received from different participants.' (FG3-E2).*

Debriefing would advise OSCE organisers and the OSCE Committee of any shortcomings or good practices of the OSCEs (Ware et al., 2014). This will allow them to take quick actions, such as removing certain items from the scoring rubric. Feedback and debriefing can help OSCE designers create better OSCEs in the future, contribute to the quality assurance process, and enhance the OSCE's catalytic effect.

The post-exam analysis includes a basic quantitative review by comparing students' scores. According to one assessment specialist, they have recently begun computing some psychometric values.

> *'Since last year, we have started off with a very basic psychometric assessment as our first step. We actually compare and contrast the scores of the students of the cohort on these six stations in between… to pick out any stations in which the students apparently performed poorly. Also, if we have two tracks, we compare the tracks' results among themselves. That, let us say, is our limited statistical analysis for our OSCE exams.' (Interview3-AS).*

However, based on some follow-up questions, it appears that the above comparison is merely comparing the results of the students to identify the outliers. They complete this without the use of any psychometric equations. As a result, this could be classified as a statistical review rather than a psychometric analysis. Likewise, they may compare other values:

> *'Comparing the students' performance in individual stations to the student performance in the whole exam and comparing the student's performance in OSCE to his performance in the course as a whole.' (FG1-D1).*

One faculty member shared his screen with me and showed me a Microsoft Excel file with a list of students in one cohort and their scores at each station. Then he explained to me how they compare the students' results. I noticed that none of the comparisons above use any psychometric calculations. However, the alpha coefficient is the only psychometric value calculated, but they did not share its results with me.

> *'We calculate the alpha coefficient that is automatically done for all the exams. But unless we go into analysing the individual checklist items within each OSCE, I believe we still have a long way with the psychometric analysis.' (Interview3-AS).*

> *'We wish to conduct full statistical analysis for each exam item (station).' (FG1-D4).*

According to the excerpt above, some participants understand the significance of conducting proper psychometric analysis. Advanced psychometric analysis can provide quality metrics and allow OSCE designers to identify some sources of errors, such as whether it is the assessor, the station design, or teaching issues (Pell et al., 2010; Tavakol and Dennick, 2011b). However, during a side conversation, I discovered they do not have any statisticians who can assist them in conducting the psychometric analysis they are aiming for.

The language as a barrier

The language used to conduct the OSCE can be a significant issue that affects its quality. This communication issue was raised by some participants as a challenging aspect.

> *'Simulated patient and student communication at times become tricky or difficult because of the language barrier. Our patients actually do not speak English, and some do not speak Arabic, but the students speak both Arabic and English. So, there is something wrong there. That could be eliminated either by an educational hospital or with the training of simulated patients.' (FG4-E5).*

Although this point was raised infrequently, based on conversations outside of FGs, it appears to have occurred with nearly all OSCEs. The language of teaching and assessment in this medical school is English. The majority of the students' and some of the faculty's first languages are Arabic. This poses two challenges, first, for the OSCE design since the OSCE designers often cannot easily find English-speaking patients or SPs. As a result, this would not allow certain OSCE stations that need some communication skills to be introduced. This may affect the validity of the OSCE.

Secondly, when the OSCE is introduced in English, most of the patients will not understand it because it is different from the community's main language. This issue questions the authenticity of the OSCE. In real life, most patients speak Arabic only, so

there may be cultural nuances important to clinical practise that will not survive translation from Arabic to English (Alnahdi et al., 2021). That is a broad authenticity issue that affects medical schools as well as hospitals in SA; however, this is not the focus of this thesis.

> *'Sometimes the language is a barrier. If we cannot find English-speaking patients, we sometimes conduct examinations in Arabic with Arabic-speaking simulators. So that is something a bit of a challenge, but also interesting to see the OSCE is conducted in Arabic, mimicking real-life conversations and then assessed obviously with an Arabic-speaking examiner.' (FG2-D2).*

Consequently, the OSCE may sometimes be administered in Arabic, but only a small percentage of examiners can speak Arabic. This would worsen the lack of examiners situation discussed later (5.3.4.1). Also, this may drive the OSCE designers to include fewer OSCE stations, which would lessen the exam's reliability, as addressed later (5.3.3.3). So here are a series of compromises which challenge the authenticity, validity, and reliability of their OSCE. Nevertheless, these skills are taught and trained in English; therefore, the assessment must reflect this. Thus, introducing the OSCEs in Arabic could be unfair to the students and lead to the loss of some examiners (Moeen-uz-zafar et al., 2015). This could compromise the students' authentic interaction with the patient during the station, affecting the results' validity (Diab et al., 2019). The participants recognise the threat this issue poses to authentic practise, but they appear unable to resolve it.

In summary, for this subtheme, the data suggest that this medical school take some steps to ensure that they implement successful OSCEs. They established an OSCE Committee to oversee the OSCE is implemented according to their standards. The key concern, however, is how the committee's staff members adopt the OSCE's standards. It is unclear whether they administer the OSCE solely based on their expertise and experience or if they embrace evidence-based practises from the wider literature. The OSCE Committee notes that they are not working alone and are considering feedback from all stakeholders. Although they strive to conduct proper psychometric analysis, they lack a psychometrician to do so. Finally, sometimes the language used during the OSCE could be a challenge for the OSCE designers, examiners, and students as this would affect the quality, authenticity, validity, and reliability of the exam.

5.3.3.2 Theory-practice gap

This section explores the mismatch between the OSCE literature and the practice of the OSCE in this medical school. I discussed three key areas where there are inconsistencies between theory and practice, i.e., between what the literature recommends and the way this institution conducts the OSCE. The three areas are only the OSCE, unobserved OSCEs, and lack of individual feedback.

Only The OSCE

The participants reported that OSCEs and work-based assessments such as mini-CEX were conducted. However, the quote below reveals that this medical school relies mainly on the OSCE as a summative performance assessment, and the mini-CEX is only used as a formative assessment.

> *'We have the standard OSCEs that take place at the campus… And then we have OSCEs in the hospital where the students go to the real environment and examine real patients… We also have a mini-CEX every week as formative assessments.' (Interview3-AS).*

Using the OSCE as the sole clinical assessment tool is contrary to the best practise of using various clinical assessment instruments within the assessment system (Khan et al., 2013b; Norcini et al., 2018). Due to the fact that no single instrument can assess all aspects of clinical competencies (van der Vleuten et al., 2010; Boursicot et al., 2020), the OSCE alone would not provide a comprehensive judgement of the students' clinical competencies. An assessment specialist suggested that they use multiple assessment instruments to improve the validity and reliability of their overall assessment outcomes.

> *'So, we know that using a variety of assessment tools increases the reliability and predictability of skills under assessment and the discrimination between the students.' (Interview3-AS).*

Although some staff are aware that this approach is not the best practise, a shortage of faculty and constrained resources may have contributed to the emergence of this situation, as discussed later (5.3.4).

Unobserved OSCE stations

Another theory-practice gap is the unobserved OSCE. Although I was initially impressed by the number of stations—twenty-six—that comprised their OSCE, on detailed questioning, it became apparent that many of the stations were unobserved.

> *'But now, we divided the OSCE into observed OSCE and unobserved OSCE. So, we have twenty stations that are unobserved… And six stations are observed that we do them on the simulators (SPs)… but both are components of the OSCE.' (FG1-D3).*

The conversation revealed that they view the computer-based questions presented to the students to be 'unobserved' OSCE stations. One participant explains the nature of these unobserved stations.

> *'A non-observed OSCE is basically a video or a sound that the student watch or listen to and writes his/her answer in the answer sheet. For example, a station with a murmur sound with some questions.' (Interview2-AS).*

According to the explanations of the participants, the students were not required to perform a skill in the unobserved OSCE; instead, they were shown a video, picture, or audio on the screen, followed by some related questions, and asked to write down the answers. Similar to knowledge test formats, this exam is administered in an exam hall

with a computer for each student. This medical school regards the above-mentioned unobserved OSCE as an OSCE exam, with each question serving as a station. The participants provided three reasons why the faculty introduced the unobserved OSCE. First, they want to enhance the reliability of their exam by introducing more OSCE stations.

> *'Naturally, you know, increasing the number of assessment moments increases reliability… So, we have tried to increase the number of OSCEs by incorporating unobserved stations into the OSCE. The OSCE committee was consistently receiving feedback that (the stations) numbers needed to go up to ten or more stations.' (Interview3-AS).*

The second reason is a lack of sufficient examiners to administer the desired number of stations, as discussed in the third theme.

> *'We do unobserved OSCE because of the same problem of constraint in the specialised examiners.' (FG2-D2).*

The third reason involved a lack of workforce and logistics.

> *'We have tried to incorporate unobserved stations into the OSCE as well, try to increase the number of stations, and overcome the lack of staff and some logistical problems.' (FG2-D5).*

The faculty appears keen to improve the OSCE's reliability and assumes that the unobserved OSCE would help increase the number of stations without the need for additional examiners or other logistics. However, according to the OSCE literature, the unobserved OSCE station, as described by the participants, is not deemed to be a proper OSCE station, as students do not perform any clinical skills (Khan et al., 2013b; Abdulghani et al., 2015; Harden et al., 2015). Although including unobserved stations improves overall reliability, it diminishes validity and authenticity since they do not meet the goal of this sort of assessment (Downing and Haladyna, 2004). After I realised this issue, I made it clear to the participants that our discussion would focus on the observed OSCE exams because I did not want the unobserved OSCE to direct the discussion, as the wider OSCE literature does not support it, and it is not the main focus of this research.

Lack of individual feedback

Personalised feedback has substantial educational value, and yet this medical school's OSCE does not provide individual feedback.

> *'I think our OSCE is not a very educational tool because there is no feedback for the students. So, how can they know what they did correctly and what they did incorrectly?… The exam itself has no built-in feedback. Feedback is necessary, so we need to work on that.' (FG3-E1).*

However, some clinical years OSCEs offer general group feedback:

> *'We give feedback to the students regarding their performance in general. We are giving it in a group, not individually. Individual written feedback could be unattainable or very difficult for the examiners to write the feedback and*

> *focus on the student's skill at the station. And also, only a few of them have problems that need feedback, so these are the ones you can select and give them feedback later on.' (FG2-D3).*

According to the above OSCE designers, individual written feedback is practically difficult. Their concern was how to implement it rather than its value. As a result, they only provide individual feedback to a few low-achieving students. However, it is more equitable and educationally beneficial if all students receive individual feedback on each station of their OSCE. Because feedback is not only about what went wrong but also about what went well, however, another participant inferred that the group feedback is adequate:

> *'I believe that our OSCE is part of the educational process because we conduct post-exam feedback for the whole cohort in which the OSCE coordinator discusses the exam with the students. This notifies the students of the strong points and weak points in their performance, ok, and puts them in the right direction. We consider this feedback as a very important learning session.' (Interview2-AS).*

The quotation implies that post-OSCE group feedback sessions have a positive educational impact; however, no evidence is provided. Although this practise is more convenient than individual feedback, it may not adhere to some good feedback principles, such as tailoring feedback to individual student performance and needs and providing specific improvement suggestions (Boud, 2015; Alomair, 2021). Furthermore, the timing of this group feedback is usually a few days after the exam, whereas effective feedback needs to be provided in a timely manner (Harden et al., 2015; Hattie and Clarke, 2018). However, the OSCE of the basic years receives no feedback at all. Therefore, some participants suggested that well-prepared cohort feedback would be beneficial:

> *'It would be a good practice if the clinical skills department gathered the common mistakes that the student attempted during the exam, then they organised a common feedback session that would be great.' (FG4-E2).*

While individual feedback is more educationally effective (Hattie and Clarke, 2018), evidence from other settings suggests that well-structured group feedback is preferable to no feedback at all (Parikh et al., 2001). Yet, the implementation of individual feedback can be more challenging. Because it necessitated a well-designed feedback system and trained examiners capable of providing written feedback immediately after the first candidate completed the station and before the next candidate began.

In summary, three issues show a gap between the OSCE practise in this medical school and the evidence-based literature. First, this medical school employs only one summative clinical assessment instrument—the OSCE—which may jeopardise the comprehensiveness of the assessment system and, ultimately, the defensibility of the school's outcomes. Second, in an attempt to increase the number of stations

implemented, this faculty sometimes uses unobserved OSCEs due to a lack of examiners, staff, and logistics. Third, the absence of individual student feedback diminishes the educational value of the OSCE. These issues suggest the need to improve these practices to align with well-established literature and enhance the quality of their OSCE.

5.3.3.3 Reliability issues

This subtheme investigates the matters that affect the OSCE reliability in this school. Respondents were concerned about certain practises that undermine the OSCE's reliability. The key issues in this subtheme are the number of stations, parallel circuits and examiner challenges, marking scheme design, and psychometric analysis.

Number of stations

The number of stations in any OSCE exam is a crucial determinant of its reliability. The OSCE of the basic years is conducted once a year, and the OSCE Committee determines the stations' number.

> *'The OSCE Committee set together to decide the topics and number of stations… We usually have from five to a maximum of six stations and rest stations in the middle. The station number depends on the examiners' availability and the ILOs of each course… The number of stations is usually five in each exam, given the limited number of examiners, but I think five is enough.' (Interview2-AS).*

This suggests that stations are decided pragmatically based on examiner availability rather than reliability and instructional alignment. A similar number of stations is conducted for the clinical years. Some participants believe that six OSCE stations are enough.

> *'Six stations for emergency medicine. They were enough.' (FG3-E1).*

According to the literature, five to six stations in an OSCE exam would not produce reliable results; thus, they argue that the minimum number of stations should be greater than twelve, taking into account the total examination time (Khan et al., 2013b; Abdulghani et al., 2015). As mentioned in blueprinting and writing the stations section, the OSCE for the clinical years in this medical school is conducted for each clinical course. This allows only a small number of skills to be blueprinted and then only a small number of stations to be implemented. However, some participants are convinced that they need to increase the number of stations.

> *'We wish to increase the number of stations. Our ambition regarding the observed station is to increase them from six stations to reach ten or twelve stations.' (FG1-D1).*

This implies that some faculty members have the knowledge and assessment literacy to highlight this, which is consistent with the "Faculty understanding" section above (5.3.2.1).

> *'We may need to increase the OSCE stations. Because at the programme review at the end of every year, we notice some of the skills students are good at and some of the skills they lack. The assessment drives learning, you know, so it could help with that. If the students know that these skills are going to be assessed, they will give them more attention.' (FG4-E5).*

This may suggest that assessment drives learning and behaviour as well; therefore, the faculty has an opportunity to align their OSCE with authentic clinical practises. Other factors preventing this faculty from implementing a greater number of stations are a lack of sufficient examiners, resources, and time, as discussed in the fourth theme (5.3.4).

> *'I think our main challenge is the number of examiners, as the number of examiners usually impacts the number of stations… If we do not have any resource and time problems, we would probably like to increase the number of observed OSCE stations' (FG1-D6).*

Regarding the stations' duration, one OSCE designer indicated that the duration of some stations might be insufficient.

> *'Sometimes, we have a short time during the assessment of each student. Some students are good students, so they can go through the station very easily. But some other students need more time to perform the skill.' (FG1-D2).*

An interviewee suggests that a content experts should observe and investigate the stations and allocate reasonable time.

> *'The specialist staff who build up this station can decide the station length. So, he may return back to the committee and in the pre-test analysis at the planning group, they can modify it, I mean this should be modified.' (FG1–D1).*

The duration of OSCE stations is an important factor that contributes to the OSCE's validity and reliability (Harden, 1990; Epstein, 2007). Stations with a duration less than that required by the task are invalid, and short total examination times may reduce exam reliability (Khan et al., 2013a).

According to the available data, the number of stations is less than what is required to produce sufficiently reliable results. Possible factors may include administering OSCE exams for short courses, a lack of examiners, and insufficient resources. This result suggests that the decision regarding the number of stations was pragmatic rather than purposeful.

Parallel circuits and examiners' challenges

The OSCE could run in one, two, or four parallel circuits in this medical school. That largely depends on the students' number and examiners' availability.

> *'The other challenge is that now we got two campuses. We have got the female campus and the male campus, and sometimes two tracks in each. The challenge is to maintain the uniformity, calibration, and inter-rater reliability of the assessment.' (Interview3-AS).*

The two-site OSCE is a widely accepted practice (Harden et al., 2015), and the assessment specialists appeared aware of some challenges for this design, such as uniformity of the students' experience and coordination of the examiners. However, there is concern about having large variations between examiners in parallel circuits.

> *'We have raised this issue so many times, but in the discussions in the post-exam reviews, we concluded that this is not an issue, as marks have not been affected as we previously thought.' (FG1-D3).*

This conclusion was reached by comparing the marks assigned by those examiners, as indicated later in this subtheme. Yet, this may only be interpreted as a rough indication of the lack of significant variation. In addition, there is no mention of whether the examiners' training helped reduce the disparity in their scores. Also, they should regulate SP performance across identical stations. Therefore, the students' performance needs to be evaluated qualitatively and quantitatively to ensure the comparability of the different sites and to account for any source of error (Pell et al., 2010; Tavakol and Dennick, 2011b), for example, by ensuring that different circuits produce comparable psychometric results. However, there is no evidence that this evaluation is performed on parallel circuits. But an assessment specialist describes a method for adjusting examiners in the following excerpt.

> *'So, we would have the same OSCE station being implemented and supervised by two or three different examiners, one in each track. And what we do is take a lot of effort to put the stations' examiners together, and then they calibrate the OSCE stations among themselves. So, they read the checklist scheme together to better understand how they can best assess the students to achieve a reliable and fair assessment to achieve high inter-rater reliability.' (Interview3-AS).*

While complete standardisation is hard to achieve due to the subjective nature of clinical skills and human behaviour, examiner training can contribute to a common understanding among examiners (Fuller et al., 2017; Schleicher et al., 2017b). Such approaches could assist in establishing a common ground for examiners and reducing extreme differences. Other OSCE examinations, on the other hand, use a single-circuit OSCE in which examiner fatigue is an issue.

> *'But the problem is that when we try to conduct more stations, there are a large number of students… We need more examiners to make two tracks, which will shorten the exam and prevent the examiner from being tired. With more examiners, we can increase the number of stations as well.' (FG3-E3).*

Examiner fatigue was frequently mentioned as an issue, as some evidence suggests that it may affect their scoring behaviour and, consequently, the OSCE's reliability (Swift et al., 2016).

> *'If we can run more than one track, it will take less time because this is for the sake of the students' accurate assessment. The shorter the time, the more attentive the examiner would be and avoid being exhausted.' (FG3-E3).*

However, having multiple circuits will necessitate additional examiners, SPs, supporting staff, and logistics, all of which are limited at this medical school, as discussed below (5.3.4). The relationship between the logistics, number of stations, total exam time, number of available examiners, examiner fatigue, and OSCE reliability is pertinent. The data from the above quotes suggest the need for additional examiners and resources, which would have significant implications for the OSCE's design and reliability. This demonstrates the tensions and compromises made by OSCE designers during the OSCE design at this school.

According to the data, this faculty conducts parallel circuits, which may affect overall assessment quality. Therefore, they attempt to ensure that the experiences of students in different circuits are comparable and that examiners are calibrated similarly. However, due to limited resources, the faculty may occasionally conduct a single OSCE circuit, which may cause examiners fatigue. Additionally, the data suggest that examiner fatigue may have an adverse effect on their scoring behaviour and, as a result, exam reliability. This suggests that the OSCE's designers could minimise this issue by improving the exam's design. This can be accomplished by having more examiners, implementing parallel circuits, and introducing frequent breaks.

Marking schemes design

The marking scheme design significantly impacts the OSCE's validity and reliability (Harden et al., 2015; Daniels and Pugh, 2018). However, an OSCE examiner revealed they were struggling with their marking scheme.

> *'A few years ago, we had problems with the checklist and the rubric that there was no sufficient information on them. But now I am a member of the OSCE Committee, and we have revised the flaws in the checklist and the rubric.' (FG4-E3).*

Although the OSCE designers stated in the quotes below that designing a scoring rubric is difficult, they believed they did that well.

> *'For the examiners' rubric and instructions, I would say, the most difficult part is, you know, designing the scoring scheme. The examiner's rubric is where we explain what zero, one, two, and three mean. So, the scoring scheme is supported by the examiner rubric and instructions in a very strong way. That is why we believe that our checklist is quite robust.' (FG2-D5).*

It appears that the faculty are considering the complexities and importance of constructing a good marking scheme. As a result, they developed a rating scale and paired it with a scoring rubric. Their marking scheme is a four-point rating scale, as described below.

> *'Which goes from poor, moderate, good, to excellent. And a final total summative mark is also given at the end.' (FG4-E5).*

It is worth noting that they considered a global rating at the end of the marking scheme, which may provide some reliability benefits (Ilgen et al., 2015). According to the OSCE designer cited below, the rating scale and examiners' instructions contributed to the OSCE's reliability.

> *'We assess the students with an excellent, very good, good, or unachieved level of performance. We can achieve good reproducibility in the exam when we give clear instructions to the examiners.' (FG1-D5).*

However, no evidence was provided in response to my request for supporting evidence, such as their psychometric analysis. But according to the following quote, the marking scheme is initially designed by one person. Then, all members of that department—the examination board—discuss the marking scheme design and how each point should be scored.

> *'I write the checklist, but I must get agreement from all the faculty in the speciality about this checklist. We look for assistance from the literature, books, and good OSCE examples from other places. We discuss the checklist in detail and how to assess the student; when to give a full mark, when to give a half mark, when not to give any mark, and what is the critical point in this skill that should be performed, which will get a higher mark than others. This discussion, I think, is very important before the exam.' (FG3-E2).*

This suggests that their design choices are influenced by literature, practise from various places, and faculty opinions. For example, they use *'critical point'* to indicate that some items in the marking scheme deserve more weight than others, which could be supported by some literature (Homer et al., 2020b). In addition, other participants indicated that the marking scheme is well-designed, allowing non-clinicians to assess the students' clinical skills in the OSCE:

> *'We try to detail the examiner instructions and marking sheet to make it easy for anyone, any examiner, even if not specialised in the (station's) speciality. We are obstetrics and gynaecology, and due to the shortage of examiners, any examiners can examine the students fairly following the instructions that we provide.' (FG2-D1).*

While the data suggests that the faculty makes significant efforts to develop good marking schemes, their design varies across OSCE exams. Some responses indicated that they use scoring rubrics with rating scales, whereas others only use rating scales, and others use checklists. This suggests a lack of consistency, which may waste faculty resources rather than focusing efforts and resources on developing a uniform marking

style. However, participants are generally pleased with their marking schemes because they enable them to employ non-clinicians and non-specialised faculty at various stations. However, no psychometric analysis has been provided to prove the quality of their marking schemes.

To summarise the findings of the third theme, the three subthemes discussed revolve around one concept: OSCE quality and design. This analysis indicates that this school places emphasis on ensuring equality, as evidenced by their establishment of an OSCE Committee that monitors the quality of OSCE implementation. However, the committee does not conduct post hoc psychometric analysis to ensure the quality of the OSCE after implementation. Furthermore, their OSCE practises indicate that there are some gaps between theory and practise, which suggests that the OSCE Committee may not be aware of these issues or may have a lack of awareness in certain aspects. According to the findings, the OSCE's use of a relatively small number of stations threatens its reliability. The analysis indicates that this school has a good OSCE design in some areas but poor practises in others; however, their focus on quality control will enable them to improve the OSCE in the future.

5.3.4 Fourth theme—Resources and infrastructure setup

This theme investigates the impact of various resources on OSCE implementation. It demonstrates the relationship between the availability of the school's resources and the actual OSCE implementation. The OSCE is resource-intensive since it demands a large number of people and logistics to implement a high-quality OSCE (Gupta et al., 2010; Patricio et al., 2013). The following statement provides a general overview of the situation at this medical school.

> *'In my opinion, all the challenges mentioned should be tackled to improve the validity of the OSCE exam. Once we have a sufficient number of examiners, staff, patients, simulated patients, tools, rooms, and scenarios, we will be able to create a valid OSCE exam.' (FG1-D1).*

According to this OSCE designer, lack of resources directly impacted OSCE implementation. Thus, this theme explores two areas: human resources and educational and logistical resources, Table 20.

Table 20. The subthemes of the fourth theme and their descriptions.

Subthemes	Description
Human resources	Examines the impact of the OSCE workforce (faculty and patients) availability and training on the OSCE implementation

Instructional and logistical resources	Explores the effect of educational activities and physical resources on the OSCE implementation

5.3.4.1 Human resources

This subtheme investigates the current state of recruitment and training of OSCE staff and patients at this medical school.

Lack of workforce

Various faculty members are involved with the OSCE, including designers, writers for the stations, examiners, and support staff. Nonetheless, the lack of examiners appears to be the biggest challenge.

> *'We are lacking manpower; the main challenge is the availability of the examiners. We prefer clinicians to be the examiners because particular stations need an expert in the subject, but usually, we do not have enough number of them, so we ask non-clinician staff to be examiners or even invite outside examiners from (name of a medical school)… This makes preparation even more complicated.' (Interview2-AS).*

Some specialised stations make locating eligible examiners more difficult. Therefore, the faculty sometimes uses non-clinicians as examiners, which can be acceptable if they assign them to the appropriate stations and provide them with adequate training (Abdulghani et al., 2015), as discussed later. This increases the need for collaboration with hospitals and other medical schools to request outside examiners.

> *'Finding and inviting them (outside examiners) to attend the exam is sometimes difficult and needs some arrangements according to their schedule in their clinical work in hospitals. This is the main challenge for this point.' (FG1-D1).*

Outside examiners can benefit OSCE implementation by covering the shortage of internal examiners and providing feedback to the OSCE designers. However, according to the last quote, recruiting outside examiners is not without complexity because outside examiners are typically busy clinicians. Therefore, OSCE designers must exert additional effort to avoid scheduling conflicts, making it difficult to find a suitable examiner. Moreover, an OSCE designer describes another challenge in the following quote:

> *'We tend to resolve this by inviting specialists and consultants from the hospitals... But this solution produces another problem because they are only familiar with clinical and hospital work and have no experience dealing with OSCE stations, how to assess students' levels, and how to run these stations rapidly to cover all (checklist) items.' (FG2-D4).*

While inviting outside examiners would bring some benefits, there is no guarantee that they will acknowledge the level of the students or the nature of the OSCE exam. The quotation implies that they may be unaware of how the students were trained to perform

the skills. Therefore, their expectations and judgements of the students' performance may not align with those of the internal examiners. Disparities in assessing attitude may compromise the OSCE's reliability, i.e., inter-rater reliability. Furthermore, the quote above assumes that hospital examiners cannot control the station's timing. This suggests that outside examiners would need some training and familiarisation with the OSCE and with the students' level (Fuller et al., 2017).

Another indication of a lack of faculty at this medical school is the scarcity of station writers and OSCE designers:

> *'But I believe that the basic science department can do better with a bit more support; you know, they may need a bigger number of OSCE designers. At the moment, I am aware that Dr (x), who is the skills lab instructor, is the only one writing and designing the stations for years one, two, and three. And my feedback as an OSCE committee chair over the years has always been that we could diversify and increase the number of designers in the basic science years.' (Interview3-AS).*

According to the assessment specialist, only one faculty member wrote and designed the basic years OSCEs. This may result in low OSCE validity and the disclosure of station information to students, as the same writer has designed the OSCE for many years. Nonetheless, this problem has persisted for a long time.

Another issue with human resources in this medical school is a lack of patients and SPs. While both real patients and SPs are limited, real patients are rarely used:

> *'As my colleagues have said many times, in fact, we do have a problem with the real patients' availability because we do not have a university hospital; I believe this is a downside. This is a major problem.' (FG4-E3).*

Since this medical school lacks an educational hospital, the majority of OSCE exams are administered on campus. Nonetheless, many of the participants argued that they required a teaching hospital to use real patients and administer the OSCE in the hospital. Furthermore, the presence of actual patients would enhance the exam's validity, whereas their absence would decrease the authenticity of the OSCE stations.

> *'Use of real patients will be better, you know, real-life scenarios with real patients in a real place will always be better. I think this could be solved if we did have an educational hospital… We may need to bring spontaneity to the interaction between the students and the patients. Because in the OSCE, everything is artificial and planned; it is not real life, not real pain, not a real situation.' (FG4-E5).*

However, due to the difficulty of having actual hospital patients, the current practice is to use cleaning staff as SPs.

> *'There are challenges, and they do form an important motivation to keep people working for improvement. One of the challenges that we have is the lack of simulators or patients… We have used the cleaning staff as simulators... We would like to have professional simulators that are highly trained.' (Interview3-AS).*

The last quote suggests that some challenges, such as a lack of SPs, can be a source of motivation for improvement (5.3.1.3). This reflects the faculty's desire for improvement discussed in the first theme. However, using untrained SPs would have a negative impact on OSCE quality and lead to the implementation of inauthentic stations (Cleland et al., 2009).

> *'I think if we keep using inexperienced simulated patients, the level of the assessment will go a little bit down. So, the simulators' level on the OSCE assessment may not give a real-life picture of the patients. It is quite different, obviously. Since we do not have professional SPs, I think using manikins could be better than using inexperienced simulated patients; this would solve some issues.' (FG4-E2).*

While manikins are also inauthentic, he preferred them over inexperienced SPs, which may indicate the difficulties in recruiting and training them.

In summary, the data suggest a lack of faculty working on the OSCE, a lack of patients, and a lack of professional SPs. For the faculty, the findings suggest that the lack of OSCE examiners, writers, and designers is a source of multiple challenges. This leads to a series of compromises, like using a low number of stations, using an unobserved OSCE, being unable to design parallel circuits, examiner fatigue, and, ultimately, lower OSCE reliability and validity than would ideally be preferred. Furthermore, recruiting outside examiners is difficult because qualified outsiders are not always available. The findings for patients noted a scarcity of real patients and the absence of professionally trained SPs. And the absence of a teaching hospital makes it difficult to acquire real patients. Consequentially, the OSCE could be inauthentic and unfeasible, as training new SPs for each OSCE would be expensive. The findings suggest that the lack of faculty and patients at this private non-profit medical school could be due to financial constraints, as mentioned in the next subtheme (5.3.4.2).

Training for stakeholders

People who contribute to the OSCE implementation need to understand how they perform their roles effectively (Fuller et al., 2017; Schleicher et al., 2017b). Therefore, this section highlights the current practice regarding the training opportunities of faculty and SPs on the OSCE exam.

The OSCE designers and examiners in this medical school came from diverse backgrounds with varying familiarity with and understanding of the nature of the OSCE, which may affect the exam's validity and reliability. Consequently, they may require training that establishes a common understanding and set of expectations for this assessment instrument. However, an OSCE examiner stated that he had no formal training.

> *'Training for the regular OSCE? No. I once received some directions from one OSCE coordinator. It was just a conversation, not proper training.' (FG4-E5).*

The academic leader endorses the last quote. He stated that a lack of assessment experts is one of the barriers impeding them from offering OSCE training. Thus, he wanted to employ some assessment experts who would be able to provide some training.

> *'If the benefits outweigh the cost, I will hire some experts who can prepare well-designed stations and train more examiners. We hope to do that soon.' (Interview1-AL).*

As a result of the lack of training, the examiners had to rely on their personal experience.

> *'I also got involved, but without any training courses. I got the experience by conducting the exam every year for the students. So, I have made it many times, but without a scientific base. Because I have not received any training, it took me so much effort and time to adapt myself to this exam.' (FG3-E3).*

This suggests that some faculty may have inadvertently pushed for an incompletely theorised OSCE, as they participate in the OSCE without a comprehensive understanding of its underlying principles. Lack of basic training may permit these examiners to manage the stations differently, thereby affecting the OSCE's validity and fairness (Schüttpelz-Brauns et al., 2019). In addition, this may influence their scoring behaviour and produce discrepancies in the marking style, consequently diminishing the exam's reliability. The field notes indicate that the examiners who reported receiving no training are relatively new to this medical school. However, I have observed that some senior faculty have received training in the past.

> *'We had regular OSCE workshops, like four or five years ago … in which we invited the faculty, and they had hands-on training on designing checklists, setting up a blueprint for the OSCE... And as expected, without calibration, the scoring was different, so this (workshops) helped us harmonise the examiners scoring behaviour.' (Interview3-AS).*

The preceding excerpt suggested that there was some training a few years ago, but it appears to have stopped or become less frequent. This is supported by the document supplied for an OSCE workshop, which has an old date. One reason could be the stoppage of the training courses provided by the partner international medical school to the faculty.

> *'We had different faculty development sessions, and most of them were conducted under the supervision of expert colleagues from (name of the partner international medical school), which is working in cooperation with our university. They checked on a lot of our activities, kind of benchmarking… which is less frequent now' (FG1-D2).*

One participant emphasised the importance of spaced training by assessment experts.

> *'I think training and learning is a continuous process. And we like to keep up with the latest developments in the academic fields. I believe training once a year or maybe every other year could be a good thing.' (FG2-D5).*

Training is also crucial for SPs and patients because they must understand their roles and how to interact and react during the exam (Collins and Harden, 1998). For example,

SPs should provide the same answers to all students in the cohort. However, the data indicates that SPs' training is brief, as it is given just prior to the exam, and thus can be considered a briefing, as described below.

> *'We usually orient them (the SPs) on the day of the exam. We usually come earlier than the students, by one hour, for example, to meet these volunteers and explain the stations to them.' (FG1-D2).*

Therefore, using inadequately trained SPs would jeopardise the OSCE's validity, reliability, and fairness (Cleland et al., 2009). An OSCE designer, however, noted that training the SPs is challenging.

> *'This is one of our challenges every year: recruiting and preparing these simulated patients to do the exam. And how can you give them clear instructions for the exam? How can they cooperate with the students during the exams? I think this makes our exam a little bit more expensive.' (FG1-D5).*

Recruitment and training for SPs may be challenging because this process must be repeated for each exam because SPs are constantly changing.

> *'If you have a simulator, you will have the challenge of training; you need to train them very well. But on every exam, you need to change them and start training again. the challenge is mostly the training.' (FG2-D3).*

According to the quote, SPs turnover from exam to exam is high, so OSCE designers should repeat the training for each OSCE exam. This could make training the SPs challenging and result in a more expensive OSCE. However, while some faculty may provide some training for SPs, certain issues may arise due to the nature of these individuals and a lack of proper instructions, raising the need for professional SPs.

> *'Here, we use the workers as simulators... So, we give instructions to our workers because some are not speaking English, are illiterate, or are not very literate, which makes training them a bit hard. But there are written instructions for the simulators, which might not be very useful for the same reason.' (FG1-D3).*

In summary, according to available data, OSCE staff training is infrequent, with junior staff receiving no appropriate training. They, therefore, work with the OSCE without any prior training. The data suggest that training used to be more frequent but has recently decreased. The findings suggest that the faculty may require assessment experts to provide continuous training and support for the OSCE implementation. Likewise, evidence suggests that current SPs training is inadequate. These challenges highlight the need for a pool of professionally trained SPs and actual patients to make OSCE implementation more feasible.

5.3.4.2 Instructional and logistical resources

This subtheme explores the availability of educational and logistical resources that can support the implementation of the OSCE at this medical school. This would assist in assessing the appropriateness of the educational and logistical infrastructure at this medical school for running the OSCE assessment.

Educational resources

This section examines the curriculum's appropriateness, teaching strategies, and learning opportunities that allow students to comprehend and perform well on this type of assessment. According to an OSCE designer, their programme allows for the progressive development of students' clinical skills.

> *'The curriculum is well organised. It is a spiral curriculum, so the students are being exposed multiple times and at different levels to the same skills.' (FG2-D2).*

Moreover, some participants highlighted that the educational system and facilities available are suitable for OSCE implementation.

> *'We have skills labs and a longitudinal clinical skills course for years one, two, and three. The skills are distributed over the whole academic year. So, the course that is going on and the skills that are taught are very closely linked. I think our educational system supports OSCE exams.' (Interview3-AS).*

> *'We have an online platform for skills in medicine, so this is a good point for our students' learning. We teach them in the college in the clinical skills lab. They practise these skills in the skill lab, and then they practice these skills in the hospital under our observation.' (FG1-D3).*

From an educational standpoint, the OSCE assessment appears viable and compatible with their curriculum and teaching activities (Hodges, 2003b). Furthermore, prior to the OSCE exams, some preparation and revision sessions are held under the supervision of medical teachers.

> *'We do a mock rehearsal OSCE exam two or three days before the actual exam to be sure every ILO is covered and the student is already aware of the process of the exam itself.' (FG1-D5).*

The practice mentioned above is akin to a formative or mock OSCE examination. While this can help students prepare for the summative OSCE, the proximity to the summative OSCE makes it exam-oriented. However, to get the most out of formative OSCEs, it may be preferable to administer them throughout the academic year but not too close to the summative exams (Pugh et al., 2018). Students can also benefit from the formative OSCE by receiving feedback on their practical skills, as there is no feedback provided during the summative OSCE (5.3.3.2).

Although the data suggest that the time allotted for students to practice clinical skills on campus may be adequate, the time allocated for skills training in an authentic environment is seen as insufficient.

> *'They do not have enough time to properly practise and master the techniques of the clinical examination in the hospital.' (FG1-D2).*

As a result, the validity of the skills examined in the OSCE may be compromised since students have not had enough exposure time to master some of the skills on which they will be assessed. This may be partially due to the absence of a teaching hospital, as discussed above (5.3.4.1).

The data suggest that this medical school's learning resources are diverse and provide information and training on clinical skills, as well as the OSCE as an assessment instrument. This would provide validity evidence that students are tested on what they have been trained on (Hodges, 2003b). However, training on clinical skills in an authentic clinical setting may not be sufficient, and the formative OSCE can be improved.

Logistical resources

In this section, I explored the physical infrastructure and resources required for OSCE implementation. It addresses logistical areas that directly impact the OSCE operation: venue, equipment, financial resources, and factors that make the OSCE logistically feasible. Implementing a reliable number of OSCE stations necessitates a large space and an abundance of equipment (Khan et al., 2013a). However, the current number of OSCE stations—six stations per exam—suggests that the campus facilities may lack sufficient space and equipment to accommodate additional stations.

> *'We have a shortage of some tools, equipment, and instruments. The availability of properly equipped places, tools, equipment, and instruments to makes it difficult to have a wide and necessary variety (of stations). Some are there, but you need a sufficient number of them to use parallel circuits, so we may need more instruments.' (FG1-D1).*

Another OSCE organiser suggested that the lack of patients necessitates the use of high-fidelity manikins.

> *'I think one of the main issues in the OSCE exam is the tools. Many tools are not available for the emergency medicine course. High-fidelity manikins that mimic real-life injuries and tools are really needed for the OSCE stations to compensate for the absence of real patients as much as possible. We now compensate for this by using video stations, unobserved OSCE.' (FG2-D4).*

Furthermore, some OSCE designers are compelled to change the topics of their stations based on the equipment at hand rather than the original blueprint.

> *'And you need to see what types of manikins and what instruments are available in your skills labs to design your station. Manikins' availability affects the station type, so the limitations of the models may restrict the freedom of choosing the most appropriate station.' (FG2-D1).*

Whereas the space and equipment available might be adequate for some OSCE exams, increasing the number of stations will require a larger space and more equipment. This also means using parallel stations is challenging. As a result, the lack of space and equipment may be hampered the implementation of a more reliable and valid OSCE (Goh et al., 2022). Some quotes suggest that financial resources may be the source of this challenge.

> *'When I attend a conference, and they propose manikins or something like that, they are usually expensive, very expensive, so we cannot afford it. And even, I mean, if you bring someone professional who can play the patient role in the OSCE, they also need special training and a good amount of money. This point is, yeah, very challenging for us.' (Interview1-AL).*

The fund, according to the academic leader, is an issue. The OSCE is a costly assessment tool (Patricio et al., 2013); therefore, medical schools should secure sufficient financial resources to enable faculty to implement a high-quality OSCE. However, the school's financial constraints prevent it from having some manikins or from providing adequate training and compensation to professional SPs. They may have other types of manikins and SPs, but the cheaper ones are usually not as good as the high-quality ones the OSCE needs.

> *'We have to pay for the simulated patients, and we have to train them as well. It is a big issue that needs a lot of work and resources. Yeah, Wallahee (swearing), we try to minimise our costs as much as we can. But, again, if you hire someone who requests a small amount of money, that usually means he is not well prepared… We also cannot bring as many outside examiners as we wish to have. But there is nothing we can do.' (Interview1-AL).*

Even though only the academic leader mentions the financial issue, he is the only participant capable of discussing it because the other participants are not involved in financial matters. However, according to the data, he appeared frustrated that financial constraints prevented them from implementing the OSCE as they desired. It has an impact on many aspects of the OSCE, such as their ability to have a bigger place, proper equipment, appropriate training, professional SPs, and outside examiners. Thus, some OSCE design decisions are based on pragmatic rather than academic considerations. Consequently, the lack of funding may jeopardise the OSCE's validity and reliability.

However, because preparing for the OSCE is a resource-intensive and time-consuming process (Barman, 2005), practices that make it more feasible and faculty-friendly are required. Participants agreed that the OSCE examination requires a significant investment of time and resources.

> *'It is time-consuming and an exhausting exam, yes.' (FG3-E3).*

> *'I think OSCE is difficult to prepare and requires a lot of assistance from different departments. Preparing for the exam is a multi-step process: to write*

> *the material, design the checklist, prepare the place and equipment, invite and train the examiners, and recruit and train the SPs.' (FG3-E1).*

Therefore, this school employs some practices to implement a logistically feasible OSCE. For instance, they have their own station bank; however, they want to expand it to make it more diverse.

> *'Regarding the scenarios, we need to have a wide range and variety of scenarios to increase the validity... We need to have a good OSCE bank with many stations and scenarios.' (FG1-D1).*

Nonetheless, the lack of faculty discussed above indicates that this medical school lacks station writers. Consequently, the use of station banks from other medical schools can make their OSCE more feasible and presumably more valid.

> *'We have one question bank from the university of (the international partner medical school's name). We also have, for example, one rotation run by (the near national medical school's name) faculty members, so we use their station bank for that course. So, these are question banks from others, yeah.' (Interview1-AL).*

The analysis of the document showed that this institution started constructing a station bank with a few stations, which can make future implementation more feasible (Abdulghani et al., 2015). They also utilise the station bank of its international medical school partner to make OSCE preparation more feasible since they do not need to create the OSCE stations from scratch.

Another step used to implement a feasible OSCE is related to the preparation for the logistics of the exam. The following example demonstrates how various units share the responsibility to assist OSCE implementation.

> *'Presently, the OSCE Committee book the room and sends the OSCE document to the Skills Lab, highlighting the logistic requirements all the way from a pen, you know, to a patient bed. ... Then the role of the Assessment Unit is to announce the exam details for the students and invite invigilators. The day before the exam, the OSCE coordinator goes to check out the preparations with the Skills Lab administrator. They go through the rooms, making sure everything is labelled and ready for the next day. I mean, it is a pleasure to walk in on the OSCE day.' (Interview3-AS).*

This suggests that different units provide important logistical support. Therefore, the OSCE designers would not be overwhelmed by OSCE logistics, allowing them to focus more on the academic aspect of the exam (Bearman et al., 2020). The organisational aspect of the exam day appears to be good, as the assessment specialist described it as '*a pleasure'*. This could be attributed to the effective preparatory phase, which supports the OSCE designers' efforts and makes the OSCE more feasible.

To summarise the findings of the fourth theme, the two subthemes examine the impact of various resources on OSCE implementation. The analysis reveals a clear shortage of faculty, SPs, and their training, which has an adverse effect on the OSCE's quality. This

indicates a lack of investment in human resources, which appears related to a lack of assessment experts (5.3.1.3 and 5.3.4.1) and funds. Moreover, the available data indicate that the present logistical resources, such as the lack of properly equipped places and equipment, pose a further hurdle to implementing sufficient OSCE stations and conducting parallel circuits (5.3.3.3). While the data suggests that the educational resources (i.e., curriculum and educational activities) are compatible with OSCE use, the participant suggests that this could be enhanced by providing students with individual feedback (5.3.3.2) and by incorporating formative OSCEs. The analysis indicates that administrative attention is required to allocate sufficient human and logistical resources to enable the faculty to implement a proper OSCE exam.

5.4 Case study B findings interpretation

This section interprets the key findings from the case study's documents, interviews, and FGs. It gives a comment on the context of this medical school. Then, I reflect on key observations that I have classified as opportunities and challenges this school encountered throughout the OSCE implementation. I conclude with an overarching theme that synthesises the concepts derived from this data analysis into a core message that can be drawn from this case study.

5.4.1 Notes on the context

Due to the fact that this is a private medical school, it is not subject to the same regulations as public medical schools. This can allow the faculty the flexibility to have their own administrative structure and embrace new educational changes. It has a partnership with a Western medical school and full accreditation from national and international educational organisations. These facts may have an impact on the institutional mission, faculty attitudes, education philosophy, and governance. Apparently, this has been represented in several areas of this institution's culture, especially its assessment culture.

Despite the fact that this institution is considered a newly established medical school, no single data point indicates that this has an impact on them. This could be interpreted as meaning that it makes no difference whether it is new or not, given that the individuals involved are not new to academia or the OSCE; they have been recruited from other places and have experience. And it is the culture of the place that they are joining which influences them and allows them to bring experience, in addition to the regulations that they were following. This may indicate that the culture of this school is supportive and receptive to change.

5.4.2 Opportunities and challenges

This section explores the main opportunities this medical school has for further OSCE implementation improvement. In addition, it will highlight the main obstacles they encounter, which may prevent them from implementing better OSCEs.

5.4.2.1 Opportunities

Although this medical school's assessment documents have some shortcomings and omissions, the availability of forms for OSCE preparation and follow-up reviews and analyses is regarded as a valuable addition. It implies that preparations for the OSCE are organised, have a proper planning phase, and are well-structured in terms of the OSCE's implementation procedure. Thus, the faculty have a good opportunity to further develop their documents to include a more detailed description of how the design decision should be made in light of the best OSCE practices.

According to the findings, the educational system, curriculum, teaching, and some facilities support OSCE implementation. Yet, the OSCE is not always adequately theorised or described in the assessment documents, and it is the only summative clinical assessment used, with no formative OSCE. This suggests that the fundamentals for a successful OSCE implementation are in place, and the school has a good opportunity to improve its clinical assessment further.

The OSCE Committee contributes significantly to their OSCE quality assurance and assists various course organisers in implementing smooth and feasible OSCEs. Nonetheless, I noticed some data points indicated a lack of assessment experts to provide consultations and training to faculty working with the OSCE. Hence, the OSCE Committee's value can be reinforced by appointing assessment experts to its board. Investing in human resources would be an asset to their OSCE and overall assessment system.

Despite data suggesting a good level of internal communication among faculty members, I observed that the responsibilities of this committee might overlap with those of other units, such as the Health Profession Education Unit and the Assessment Unit. For example, this may fragment the work and ultimately potentially affect the OSCE's smooth operation. Therefore, restructuring these units may help avoid unnecessary conflict in responsibilities while being more cost-effective.

According to the data, the faculty initiated a good but limited collaboration with hospitals, other local medical schools, and the international partner medical school. Thus, there is an opportunity to take this collaboration to the next level to facilitate sharing experiences, resources, examiners, and SPs. For example, the proposal for establishing an Assessment and Examination Centre. This will assist in enhancing the school's OSCE feasibility and overall implementation quality, e.g., by providing enough SPs and

manikins, especially since a teaching hospital is not available. However, the school appears to lack the capacity and resources necessary for this project.

Multiple data indicate that this school's staff, who come from diverse backgrounds, are enthusiastic and have a good understanding of the OSCE. And this medical school's culture is welcoming and flexible to change. Due to this, the faculty is enabled to bring some experiences and practises from other places. This is a valuable opportunity for this school to continue improving their OSCE. Even though the findings indicate that the OSCE is not perfectly implemented, several participants indicated that it still offers educational benefits. This can be interpreted as an indication that the faculty is on the right track, which can be a motivator to invest more to maximise the OSCE benefits and promote students' future practice.

5.4.2.2 Challenges

According to the findings, the main challenge this medical school faces is a lack of resources available for the OSCE. There are insufficient human resources, including assessment experts, OSCE writers, examiners, and patients/SPs. There is also evidence of inadequate training for examiners and SPs. In addition, some logistical deficiencies are evident, including a lack of equipment and space, the absence of a teaching hospital, and financial constraints. The aforementioned resources are crucial for implementing a high-quality OSCE. Consequently, the lack of resources renders their OSCE design pragmatic, and the selection of station topics is greatly affected by the availability of resources. Several data points, for instance, indicate that the number of stations is contingent on the availability of examiners. The data reveal that the number of stations is suboptimal and may not be able to produce a reliable result. Another example is that some stations are designed based on equipment availability rather than solely on blueprinting results. Therefore, if the school did not invest more in these resources, it would be unable to improve its OSCE practice.

In an attempt to overcome this challenge, the faculty introduces an unobserved OSCE, which is less resource intensive. Nonetheless, as discussed previously, this cannot be considered an OSCE, which also reflects a theory-practice gap at this school. Instead, this school can expand its collaboration with other institutions to reduce costs while increasing benefits by exchanging available resources. Moreover, a more integrated OSCE approach may allow the blueprinting of more skills, allowing adequate OSCE stations to be implemented to produce a reliable result. For instance, instead of conducting one OSCE for each course, use a single, longer OSCE examination each year so that all skills taught that year could be blueprinted.

Although some quality control measures are implemented following the OSCEs, a proper psychometric analysis is not carried out. Station-level psychometric analysis can provide useful information, such as error source identification and setting the standard (Pell et

al., 2010; Fuller et al., 2013). However, the lack of an experienced psychometrician makes this analysis impossible within the current staffing, which can also be attributed to a lack of human resources.

Following the exam, the feedback for the students is provided to the cohort, but no individual feedback is provided. Some participants stated that there is no built-in method for providing individual feedback for each student and that attempting to do so is difficult, given the scarcity of examiners.

5.4.3 Overarching theme: A series of dilemmas and compromises

This overarching theme is a concept clustering for the findings analysed in this case study, which provides insights into relationships between concepts generated by data analysis. This would allow a better comprehension of a complex operational and educational process involving multiple stakeholders and variables. A reflection across the four themes of this case study reveals that it is all about a series of tensions and dilemmas. There are multiple design compromises in this school's OSCE implementation. For example, examiners, SPs, training, funds, and space all push faculty to settle on the middle ground by making hard choices to implement the OSCE. While some of these dilemmas and tensions are understood and articulated by faculty, they appear not to comprehend all of them.

According to the document analysis, the school is keen to have a high-quality assessment. In contrast, the in-depth analysis of the interviews and FGs reveals that the school is unable to implement an OSCE that meets the requirements stated in the documents. The faculty often makes compromises without necessarily realising their interconnection, resulting in a conundrum at all levels. Although participants are aware of many of OSCE's best practises, contextual obstacles make it difficult for them to completely adhere to them, resulting in compromises.

The findings suggest that, based on the available resources, the OSCE implemented at this medical school does some of the basic work well. At face value, these efforts seem important and valued. But in fact, the school may be unable to improve the OSCE unless it invests in faculty and resources, as this is the source of many shortcomings in their OSCE practice. Furthermore, the faculty should consider restructuring the OSCE within the assessment system to embrace innovation, capitalise on opportunities, and address challenges to implement high-quality OSCEs.

5.5 Summary

This chapter presented the findings with analysis from case study B documents, interviews, and FGs. Using Codebook analysis, the findings of the school's assessment documents revealed that they cover most of the components of good assessment described in the literature but with little theoretical description and some omissions in places. On the other hand, the findings from the interviews and FGs were generated using RTA, which produces perhaps more critical findings. I composed the findings into four key themes: institutional and assessment culture, faculty expertise and practices, OSCE quality and design, and resource and infrastructure setup. The analysis and interpretations indicated that the current OSCE practise at this medical school has both positive and negative aspects, and I highlighted the factors that affect its implementation. The following chapter will bring the findings from the two case studies together and relate

them to the existing literature in order to answer the research questions and highlight the contribution of this research to the literature.

Chapter 6 Discussion

6.1 Introduction

This chapter presents a synthesis and discussion of the major findings of the two case studies and relates them to the existing literature in order to answer the two questions of this research. Firstly, a consideration of how the OSCE is being implemented in Saudi medical schools. Secondly, a summary of the challenges and opportunities offered by adopting the OSCE in these schools. Then, I conclude by highlighting the overarching theme, differences between public and private medical schools in OSCE implementation, and key messages for medical schools to improve their OSCE implementation.

6.2 The OSCE implementation in Saudi medical schools

This section brings the two case studies together to compare, contrast, and discuss commonalities and uniquities. It offers a synthesis and discussion of overall findings to answer the first research question, “How is the OSCE being implemented in Saudi medical schools?”. The findings indicate that each case study's OSCE implementation practises are different to some extent. According to the analysis, school B appears to have a better overall implementation of the OSCE than school A. The findings provide a documentary analysis and four themes across the two case studies, which I will discuss in this section to demonstrate how they relate to the literature and what this means for OSCE practise.

6.2.1 Documentary analysis discussion

This section discusses the main findings from the documentary analysis of the two case studies. According to Downing (2006) and Johnson et al. (2008) assessment documents are an essential element in establishing assessments fundamental decisions, rules, regulations, specifications, and design processes. I compared the case studies assessment documents findings with three international organisations in medical education (Appendix 17). The three organisations whose criteria and frameworks I utilised are the Ottawa Conference, the AMEE ASPIRE-to-excellence award in assessment, and the WFME (Boursicot et al., 2011; WFME, 2015; Boursicot et al., 2020; WFME, 2020; AMEE, 2023). I utilised these resources because they are well-known for promoting excellence in medical education assessment.

Although the codes that emerged from the Codebook analysis of the assessment documents of the three organisations in relation to the case studies documents are similar, there are multiple omissions in the two case studies (Appendix 17). According to the evidence, the majority of the criteria for good assessment are mentioned in both case studies; however, the evidence also suggests that the OSCE lacks a connection to the

broader assessment system. For example, there is no mention of the need for coherent and feasible assessments, nor are the processes for the quality control measures for the assessment instruments addressed in detail. Furthermore, there is ambiguity regarding which competences should be examined. This suggests the need for more comprehensive and detailed assessment documents to guide a successful OSCE implementation.

In the two case studies, the OSCE is used as the sole clinical assessment instrument from the first to the fifth year. The purpose of using the OSCE in these medical schools as a high-stakes summative exam is obvious to many faculty members; however, other stakeholders, such as new faculty, employers, regulatory bodies, and the public, may be unaware of the importance of the format. There is no explicit description of the OSCE's purpose, how it integrates into the wider assessment system or indication of the need for multiple clinical assessment instruments. As stated in the documents and discussed in sections (4.3.3.2 and 5.3.3.2), overreliance on one clinical assessment instrument may jeopardise the validity and reliability of their overall clinical assessment system (Shumway and Harden, 2003; Norcini et al., 2018). Thus, any shortcomings in their OSCE would undermine the credibility and defensibility of their overall outcomes (Colbert-Getz and Shea, 2021). It is widely argued that a single clinical assessment cannot adequately judge students' clinical performance (Epstein, 2007; Khan et al., 2013b). For example, OSCE lacks some authenticity elements because it is conducted in artificial settings (Boursicot et al., 2020; Flood et al., 2022; Hyde et al., 2022). Therefore, medical schools may need to include other performance assessment instruments, such as Mini-CEX, DOPS, or other WBAs, as each has advantages and limitations (van der Vleuten et al., 2010).

Harden and Gleeson (1979) created the OSCE primarily to assess students' clinical skills. Furthermore, because the OSCE is an expensive and complicated assessment to design, the school should make the best use of its resources. However, the lack of a clear statement determining the use of the OSCE may help explain why medical school B (5.3.3.2) incorporates unobserved OSCE stations while not assessing any of the students' clinical competencies, thereby wasting valuable school resources. At both medical schools, there is a lack of clear statements defining the intended use of the OSCE to support the decisions made during all phases of OSCE implementation, which will benefit the school in avoiding shortcomings and managing expenses.

The documents of both medical schools indicate that the OSCE pass-score is fixed of sixty per cent to be in line with their universities' regulations. While this approach would be regarded as comfortable and cost-effective for the schools in the absence of a professional psychometrician, this will undermine the defensibility and credibility of the OSCE results (Pell and Roberts, 2006; Tekian and Norcini, 2015). Other approaches would produce more defensible results for high-stakes exams. Kramer et al. (2003),

Yousuf et al. (2015), Dwivedi et al. (2020), and Homer et al. (2020a) recommend using criterion-referenced methods. They argue that the criterion-referencing approach of the Borderline Regression Method (BRM) for standard setting is deemed fair and defendable for high-stakes OSCEs. Because a summative OSCE should produce just and defensible results, it must be grounded in evidence-based practices. Using guidance from assessment regulatory documents and assessment specialists should make an informed decision between criterion and norm-referencing methods when selecting the appropriate standard setting approach. Their decision should not be arbitrary but logically justifiable considering the context (Homer and Darling, 2016), as discussed in the literature review chapter.

6.2.2 Interviews and FGs discussion

While the themes generated in both case studies are similar, several findings are specific to their respective contexts. Table 21 illustrates the key difference in OSCE implementation between case studies A and B, which are structured according to the themes. The entire contents of Table 21 are drawn from the Findings and Analysis Chapters (Chapter 4 and Chapter 5). I will discuss the role of culture, faculty, design, and resources in the OSCE implementation, highlighting the commonalities and differences. Then, I will discuss the overarching theme of both case studies: a series of OSCE implementation dilemmas and compromises.

Table 21. Key differences and similarities between case studies A and B.

Theme	Area	Medical school A	Medical school B
Institutional and assessment culture	Funding source	Public sector	Private sector
	National accreditation	Partial accreditation	Full accreditation
	International accreditation	No	Yes - by two agencies
	Collaboration with an international medical school	No	Yes
	The documents and policies' comprehensiveness	-Comprehensive -Lack of procedural documents (Templates and forms)	-Not comprehensive -Focusing on procedural document (Many templates and forms)
	Is passing the OSCE required to pass the clinical course?	The school has not reached a clear decision in this regard	No – only 60% of total assessments in any course were necessary to pass that course
Faculty expertise and practices	Inexperienced or new faculty	Many	Few
	Nationality of the staff	The majority are home faculty	The majority are international faculty
	Faculty understanding of the pros and cons of the OSCE	Well-informed	Well-informed
	Preparing for the OSCE	Few days (5-10 days) ahead	12 to 2 months ahead Everything should be ready one month before the exam
	Assign examiners to the stations and send material to them	30 minutes before the exam	2 - 3 days before the exam
	Briefing prior to the OSCE	No	Yes, but generic
OSCE quality and design	OSCE Committee availability	No	Yes
	The similarity of OSCE design and implementation across various school courses	Variable	Very similar
	The average number of stations in different courses	2 - 5 station	5 - 6 stations
	Stations' duration	Ranged from 3 to 15 minutes	10 minutes, or longer if necessary
	Feedback to students	No individual feedback	No individual feedback
	Marking schemes design	Different from exam to exam	Using a unified form, but it may vary
	Post hoc statistical analysis	No	Yes, but very basic
Resources and infrastructure setup	Training for OSCE examiners	No	Yes - intermittently
	Training for outside examiners	Not mentioned	Not mentioned
	Training for SPs	Yes, but limited	Yes, but limited
	Availability of the examiners	Very limited	Very limited
	Availability of real patients	Little	No availability
	Availability of SPs	Low availability and usually, inexperienced SPs	Low availability and usually, inexperienced SPs
	Venue and equipment	Limited	Limited
	Funds	Financial constraints	Financial constraints
	The need for an educational hospital	Yes	Yes

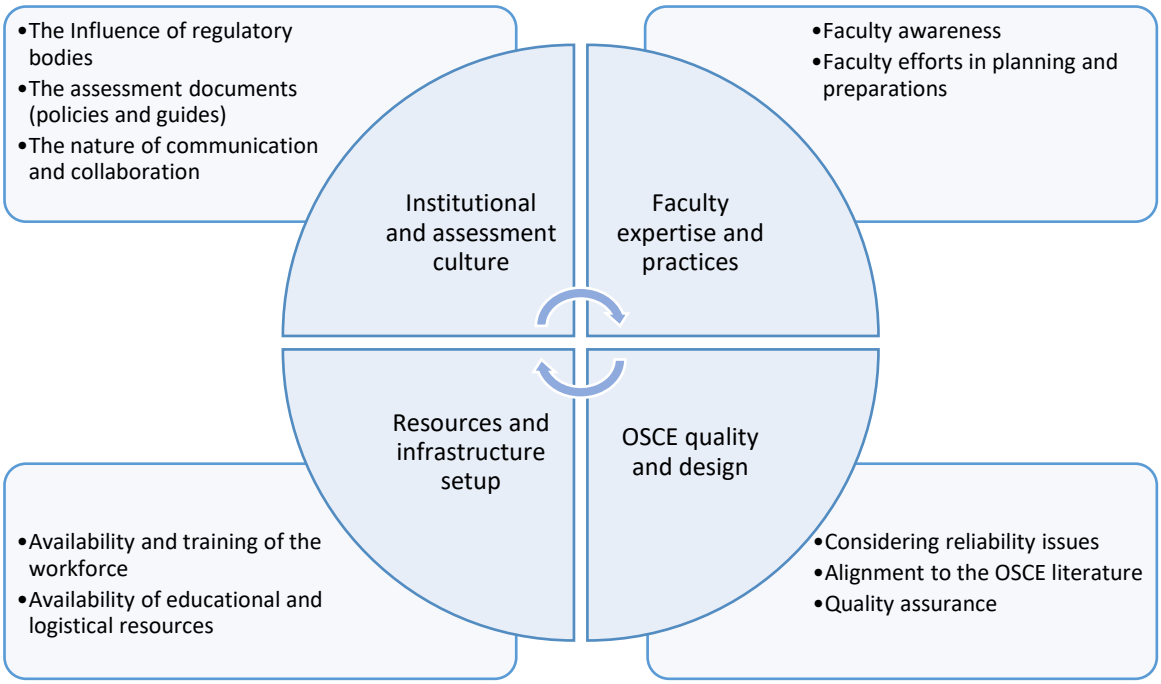


Figure 15. This diagram summarises the key findings from the two case studies. This conceptual framework illustrates how these components interact throughout the OSCE implementation process.

The diagram above illustrates the four themes that directly affect the implementation of the OSCE (Figure 15). The analysis revealed that a good OSCE requires integrating the four components in the diagram and careful consideration of their constituent elements, all of which are required to implement a high-quality OSCE. Therefore, assessment is more than simply ensuring the technical and practical aspects of the exam (i.e., design and resources) without careful consideration for wider influencing factors (i.e., institutional culture and faculty expertise). While much of the literature focuses on the practical steps to implementing the OSCE (Gormley, 2011; Zayyan, 2011; Abdulghani et al., 2015; Daniels and Pugh, 2018), which are essential, this study takes a more holistic approach by considering the broader influencing factors.

6.2.2.1 First theme: Institutional and assessment culture

Case study A is a public-sector medical school, while case study B is a private-sector medical school. The funding source appears to influence how the school operates in general and approaches the OSCE examination in particular, which aligns with Altbach and Levy's (2005) conclusions about private higher education. Although Case Study B's business model is a non-profit private medical school, my data interpretation indicates that it runs with a private sector attitude. This includes profit-oriented decision-making and an eagerness to deliver high-quality education to rival other medical schools and

keep enrolling good students. For example, one participant's suggestion for OSCE improvement was to establish an Assessment and Examination Centre, and his comment was: ‘so we can get more assessment experts and OSCE examiners as well as making profits.’ Case study B (Interview2-AS). However, the data analysis indicated that private school (B) shows an overall better performance in the OSCE implementation than public school (A). This finding is consistent with a Mexican study that indicates a higher quality education in the country’s private medical schools in comparison to the public medical schools, which reflected a better performance for the private school graduates in the national AKT test in comparison with those from the public schools (Hernandez-Galvez and Roldan-Valadez, 2019). However, Joy et al. (2007) find that the Indian community prefers doctors who graduated from public medical schools over those who graduated from private ones, indicating that the community is doubtful about the quality of private schools. Although no published research has compared the assessment quality of public and private medical schools, based on my limited data, I inferred that some private schools might be more successful in implementing high-quality OSCE. This could be due to their greater agility to improve their practice, stronger decision-making autonomy, and desire to maintain a good reputation, as indicated in Table 22 and Section 6.4.2. However, since this issue is not the focus of my thesis, I have recommended it as a potential area for future research to investigate (8.4.3), for example, whether this is influenced by how public and private schools are regulated and whether this difference is typical. Furthermore, the available data suggested that school B seemed more agile in making decisions as their leader showed the ability to make autonomous decisions (5.3.4.1, p183) compared to the leader of public school (A), who has less autonomy, probably due to the organisational structure and hierarchy of the public medical schools (4.3.4.2, p130).

Despite the fact that both medical schools are relatively newly established and have submitted applications to the national accreditation agency, only medical school B has obtained programmatic accreditation from the NCAAA. Additionally, school B has international accreditation. This suggests that this private school has more reasons, as stated above, to control assessment quality and obtain accreditation than the public-funded school, which may have a different motivation to control quality, possibly merely for accreditation purposes. Weiner (2009) and Shiffer et al. (2019) argued that the regulatory bodies (i.e., accreditation agencies) are influencing the institutional culture and assessment quality. The findings of these papers are consistent with the findings of this study, which indicate that school B has a better OSCE implementation compared to school A. This study suggests that school B adheres to some rules established by the regulatory agencies, indicating that the OSCEs are held to the same high standards. However, this study's findings suggest that the accreditation authority is not as influential as some authors suggest (Sanyal and Martin, 2007), but it still has some favourable

effects. For example, school B obtained two accreditations; however, their OSCE practice is not as described in the literature as good OSCE practices, and school A has a poorer OSCE implementation and no accreditation. This finding is consistent with the notion that accreditation alone is insufficient to ensure the quality of a school's assessment, especially with the lack of follow-up after getting the accreditation (Harvey, 2004). Often, it can only partially contribute to the assessment quality.

Another distinguishing feature of the two case studies' institutional cultures is the nature of internal communication between the faculty of each case study. According to the data, the communication and information delivery level is more effective in school B than in school A. These findings are consistent with the literature showing that effective communication among faculty members is crucial for the success of the educational process (Thornhill et al., 1996). Banihashemi (2011) and Suter et al. (2009) emphasise that good dissemination of rules and regulations and accessibility of feedback channels are essential for a healthy institutional culture. This explains the impact of ineffective communication on the OSCE implementation in school A. For instance, section 4.3.1.2 shows that the assessment specialists were uncertain whether the passing score had been changed from seventy to sixty and whether students who failed the OSCE could pass the course or not. Such issues highlighted two main challenges in school A, which are a lack of effective communication and a lack of clear regulations.

The lack of clear internal regulations is evident at school A; consequently, the staff working with the OSCE tend to rely on their personal experience without support from other departments (4.3.1.2). On the other hand, school B has the OSCE committee to monitor and support its OSCE implementation (5.3.3.1). These findings align with the need for supportive assessment guidelines recommended by Simper et al. (2022). Furthermore, the data suggest that collaboration with external parties is more evident in school B than in school A (4.3.1.1 and 5.3.1.1). The data indicate that school B collaborates with some national medical schools, one international medical school, and hospitals, which makes their OSCE implementation more feasible and probably more valid by including external experiences.

An important finding and positive cultural aspect of both medical schools is the strong desire for improvement among all participants. This section and the sections below highlight several challenges to overcome and opportunities to utilise to improve the OSCE implementation. Therefore, those enthusiastic faculty members can find plenty of room for improvement (Alsharif, 2011) as both medical schools are newly established, so they are not bound by history. Nonetheless, the data suggests that the private medical school (B) can be more responsive to introducing a change because it has better internal communication and external collaboration (5.3.1.1) and is more flexible and adaptable to change (5.3.1.3). Although both medical schools have the opportunity to change and improve their OSCE implementation to align with the government's 2030 vision to

enhance higher education, the private school may have a higher chance of success due to the privatisation movement in SA supported by the Saudi Vision 2030 (Vision2030, 2023). The institutional and assessment cultural aspects discussed in this section lead to a different assessment product; therefore, the schools need to consider them to implement a high-quality OSCE.

6.2.2.2 Second theme: Faculty expertise and practices

The staff members' experience and understanding of the OSCE's purpose and structure will enable them to administer exams effectively (Berendonk et al., 2013; Bearman et al., 2020). This study suggests that the participants' theoretical understanding of the principles of the OSCE between the two case studies is comparable. However, the staff of case study B often appear more thoughtful and critical regarding the OSCE. For example, in section 5.3.2.1, one participant commented: *'The OSCE needs a good design to be effective.'* case study B (Interview1-AL) suggests that the actual design is more important than mere theoretical understanding. The findings indicate that the OSCE design in school B is superior to that of school A, which can be largely attributed to the large number of new staff in school A. Case study A faculty members are typically young clinicians who come from similar backgrounds with limited experience in medical education. In contrast, the faculty members of case study B are more senior, come from diverse backgrounds, and have greater academic expertise; thus, they have previously gained some experience that they brought with them. This suggests that schools with experienced faculty members are better equipped to implement better OSCEs, which is consistent with the findings of Bearman et al. (2017).

The results indicate that participants from school B are more satisfied with their OSCE implementation (5.3.2.1) than those from school A (4.3.2.1). However, all participants' opinions agree that the OSCE is a valuable assessment instrument. This finding is consistent with Alsaid and Al-Sheikh (2017), Majumder et al. (2019), and Solà-Pola et al. (2020), who all concluded that the OSCE is a useful, fair, and acceptable assessment instrument. Although the participants from the two medical schools convey that their OSCEs are suboptimal and require enhancement, they affirm that the OSCE positively impacts the clinical practise of the students after graduation (4.3.2.1and 5.3.2.1). While one published work does not support this result (Rudland et al., 2008), the conclusions of the majority of studies are aligned with the findings of the two case studies (Graham et al., 2013; Komasawa et al., 2020; Horita et al., 2021). This is a significant finding that can be attributed to the fact that the students' OSCE experience in medical schools prepares them for comparable examinations in post-graduate training. However, this result should be taken with caution due to the possibility of social desirability bias (i.e., participants favouring their outcomes)(Grimm, 2010); and a lack of third-party benchmarking (i.e., no external comparison for performance) (Engelkemeyer, 1998).

While both faculties are exerting efforts to develop the OSCE, evidence suggests that the faculty of medical school B has a better understanding of and capacity to manage the OSCE's implementation. The OSCE literature emphasises that the OSCE is a complex and resource-intensive assessment that necessitates a thorough preparation phase and trained faculty (Boursicot, 2010; Khan et al., 2013a; Harden et al., 2015; Bearman et al., 2020). In accordance with the literature recommendations, school B starts the planning and preparation phase several months before the exam (5.3.2.2), whereas the staff in school A do so just days beforehand (4.3.2.2). This implies that some staff members in school A underestimate the complexity and difficulty of implementing an effective and high-quality OSCE, resulting in an overall substandard implementation. For example, school A does not always allow sufficient time to carefully write and review the OSCE stations (4.3.2.2), and no participant in either case study indicated that they piloted the station beforehand, as the literature suggests (Abdulghani et al., 2015; Harden et al., 2015).

According to Newble (1988) and Raymond and Grande (2019), creating a blueprint is a critical step in planning the OSCE, which is an important source of validity to ensure that the exam accurately reflects the taught skills. According to the data, staff in medical school B are keen to blueprint all their OSCEs, whereas OSCE examinations at medical school A are typically not blueprinted (4.3.2.2 and 5.3.2.2). Another validity aspect raised by many examiners is the importance of briefing them and sending the OSCE materials to them before the exam day. The OSCE literature emphasises the role of examiner briefing in conducting a valid OSCE (Harden, 1990; Fuller et al., 2017; Gilani et al., 2021; Malau-Aduli et al., 2023). However, data from school A indicated the absence of briefing (4.3.2.24.3.4.1), whereas school B examiners indicated that they received a briefing and the OSCE material a few days prior to the exam (5.3.2.2). They suggest that this step allows them to comprehend the OSCE station details and provide feedback on the OSCE design to the OSCE designers. Whereas school B shows practices supported by the literature and enhances the OSCE's validity, school A fails to establish many of these practices.

6.2.2.3 Third theme: OSCE quality and design

The analysis of the findings indicates that case study B takes more measures to ensure the quality of its OSCE design than case study A. While data shows that both medical schools are keen to design a good OSCE, medical school B established an OSCE Committee as a quality assurance measure. This committee coordinates, supervises, and provides support for the OSCE implemented across the school, which enhances the quality of and staff satisfaction with their OSCEs (5.3.3.1). However, excessive supervision would discourage innovation and creativity in designing the OSCE; thus, the OSCE committee should consider this issue to allow the OSCE to evolve in accordance

with the context (Shand, 2020; Chan et al., 2023). In contrast, the OSCEs in case study A are designed and administered by departments without evidence from the data indicating any input from assessment specialists or the exam committee, allowing for variations in OSCE design without knowing its quality (4.3.3.1). School B pre-exam review follows the good OSCE practices in OSCE implementation (Abdulghani et al., 2015; Harden et al., 2015; Schuwirth et al., 2021). The literature suggests that pre-exam analysis for the stations will minimise their flaws and ensure the validity of their content.

The post-exam analysis is also important, and the finding indicates that both case studies receive feedback from various stakeholders. According to Pugh and Smee (2013) and Ware et al. (2014), feedback from examiners, students, and SPs to the OSCE organisers is a valuable source for quality assurance and improvement. In school B, the findings indicate that this feedback is considered to inform future OSCE and curriculum design (5.3.3.1). However, the participants from case study A show less emphasis on the feedback role in their school and OSCEs (4.3.3.1). According to the literature, maintaining open feedback channels between stakeholders and faculty allows for a culture of continuous improvement and demonstrates the faculty's willingness to improve (Stassen et al., 2004). As a result, school B interviewees believe that their OSCE implementation is improving over time, leading to more accurate decisions regarding students' competencies.

Psychometric analysis is an effective approach to evaluating the quality of OSCE design outcomes and identifying station-level issues (Newble and Swanson, 1988; Pell et al., 2010; Tavakol and Dennick, 2011b). The literature recommends employing a combination of metrics, such as Cronbach's alpha, coefficient of determination (R^2), between-group variation, and the number of failures, because no single metric alone is sufficient to provide a comprehensive picture of OSCE quality. Although case study B demonstrates that some statistical values are evaluated, e.g., Interview3-AS said: '*If we have two tracks, we compare the tracks results among themselves. That, let us say, is our limited statistical analysis for our OSCE exams*.', neither of the case studies conducts a sophisticated post hoc psychometric analysis of the OSCE outcomes (4.3.3.1 and 5.3.3.1). Psychometric analysis is important to ensure exam reliability, identify the source of error, and ensure the quality and defensibility of OSCE results (Pell et al., 2010). It is also useful to triangulate local results with other well-established exams, licencing examinations, and hospital clinical performance (Boursicot et al., 2020). However, according to case study A and B participants, it is not being conducted at either school due to a lack of qualified psychometricians (4.3.3.1 and 5.3.1.3).

In accordance with a study about the language barrier between the students and the SPs in OSCE in SA (Moeen-uz-zafar et al., 2015), both case studies mention the same concern. All most all students are Arabs who were taught and assessed in English, but according to the participants (4.3.3.1 and 5.3.3.1), SPs typically do not speak English.

This could compromise the quality of communication during the station and, subsequently, the validity and authenticity of the examination (Mirza and Hashim, 2010; Diab et al., 2019). According to Al-Mously et al. (2013) and Mann et al. (2013), introducing some activities in English has a negative impact on the academic performance of students for whom English is a second language. This is a broad issue of authenticity because the majority of patients in hospitals are Arabic, and the students prefer to interact in Arabic, but they were instructed and assessed in English (Alnahdi et al., 2021; Alomair, 2021).

The clinical competencies being tested are indicated in the assessment documents of school A, which include history taking, physical examination, simple procedures, patient management problems, communication, and attitude 4.2.2.2); and school B indicate that they adhere to the SaudiMEDs framework for clinical competencies (5.2.2.2). Although many OSCE stations in both schools assess real clinical competencies, there are many other examples where the OSCE stations do not include any clinical competencies. School A, for example, uses 'written OSCEs' (4.3.3.2), whereas School B uses 'unobserved OSCEs' (5.3.3.2). This can be attributed to a lack of clarity regarding the purpose of this assessment tool and what competencies should be tested. Although adding written and unobserved stations can improve the OSCEs' overall reliability, it reduces their validity and authenticity as they do not fulfil the aim of this type of assessment, i.e., students are not observed performing a skill (Downing and Haladyna, 2004; Ware et al., 2014). There are more effective and feasible methods to assess the elements assessed in these written and unobserved stations, such as through written exams that include pure data interpretation (Amin et al., 2006).

In some situations, there is a theory-practice gap in the design of the OSCE. In both case studies, faculties administer one OSCE exam for each clinical course, with an average station count between two and six (4.3.3.3 and 5.3.3.3). This is a small number of stations that are deemed unlikely to produce valid and reliable results (Swanson et al., 1999; Abdulghani et al., 2015; Harden et al., 2015). According to the data analysis, this is primarily due to four factors: some courses only have a limited number of performance skills to be assessed (4.3.3.3), there is a lack of examiners, a lack of SPs, and a lack of equipment and space (4.3.4 and 5.3.4). This highlights the importance of considering the purpose of this assessment, i.e., summative or formative (Boursicot et al., 2022). Then, the schools must address these issues and may need to reconsider their assessment system to implement an OSCE that covers multiple modules so that sufficient skills can be blueprinted (Pugh and Smee, 2013; Yudkowsky, 2019). Running one large end-of-year OSCE will also be more feasible from an organisational and financial standpoint. In addition, the administration has to invest more in human and logistical resources to ensure the successful implementation of OSCEs, as discussed below (6.2.2.4). These challenges are why case study B faculty use unobserved OSCEs to increase the number

of stations (5.3.3.2). Similarly, due to a lack of examiners and SPs, school A uses 'written OSCE' (4.3.3.2), which, as previously discussed, is an inappropriate substitute for the real OSCE.

Another theory-practice gap in the two case studies OSCEs is the lack of individual feedback for students. Although the OSCE designers in school B may provide post-exam feedback to students, this is done as a group rather than individually (5.3.3.2). Case study A data indicate that their OSCE does not have a system for providing students with feedback (4.3.3.2). According to Ossenberg et al. (2019), Davies (2023), and Boud (2015), effective feedback must be timely, specific to the student's performance, and include suggestions for improvement. Cohort feedback does not appear to meet any of these criteria (Hattie and Clarke, 2018; Dawson et al., 2019). The aforementioned theory-practice gap issues highlight the need for schools to align their OSCE practises with the evidence-based practises described in the literature to conduct high-quality OSCEs.

Findings show that both case studies use checklists and rubric marking schemes, and in school B, they may also use global rating. In their systematic review, Ilgen et al. (2015) concluded that while checklists may have higher inter-rater reliability, global ratings have higher inter-item and inter-station reliability. However, some authors argue that the checklist is better for inexperienced examiners assessing procedural skills (Ma et al., 2012), whereas others contend that the global rating is superior for expert examiners (Regehr et al., 1998). Nonetheless, other authors believe that if carefully structured, all types of marking schemes would yield reliable and valid results (Daniels and Pugh, 2018; Wood and Pugh, 2019). Both case studies have templates for both types, but not all OSCE designers use them. Typically, they use either detailed checklists or a four-point Likert scale with written marking criteria. Participants from School B were satisfied with their marking scheme design and deemed it suitable for use by non-clinical examiners (5.3.3.3). In school A, however, some examiners were pleased with the checklist with differentially weighted items that they designed in the department, as they said it was more credible and minimised examiner variation (4.3.3.3), which is supported by some literature (Homer et al., 2020b). But another examiner in school A indicated that the checklist used in his department is not well-designed and it was a source of variation between the examiners. The data indicated that the design of the marking schemes differs between schools and departments, with various degrees of examiner satisfaction about their quality. However, it requires to be carefully designed to enhance the validity and reliability of the OSCE (Dyrvig et al., 2014; Malau-Aduli et al., 2021).

6.2.2.4 Fourth theme: Resources and infrastructure setup

According to studies on the challenges in Saudi medical schools, one of the main challenges for Saudi medical education is a lack of faculty (Hamdy et al., 2010; Tekian and Almazrooa, 2011). In accordance with those results, the data from the two case

studies reveal a workforce shortage, including OSCE examiners, station writers, and assessment experts (4.3.4.1 and 5.3.4.1). Nonetheless, the documentary analysis of case study A states the need for two independent examiners in each OSCE station, as suggested by Brannick et al. (2011), whereas the majority of the psychometric literature recommends using one examiner but with a larger number of stations and emphasises the importance of proper training (Newble and Swanson, 1988; van der Vleuten, 1996; Besar et al., 2012; Abdulghani et al., 2015; Harden et al., 2015; Faherty et al., 2020). This is an excellent example of how, in some instances, the absence of alignment with evidence-based practices leads to an undue burden on logistics.

In line with the recommendation of Abdulghani et al. (2015), both medical schools utilise outside examiners to compensate for the shortage of internal examiners. However, this does not appear to effectively address the issue because these outside examiners may not always be available in sufficient numbers. Moreover, the findings suggest that because outside examiners do not receive appropriate training, there is no assurance that they will recognise the students' level or the format of the OSCE examination (4.3.4.1 and 5.3.4.1). Therefore, they may have different marking behaviours and manage the stations differently, which will adversely impact the OSCE's validity and reliability.

Similarly, the finding indicates a lack of internal examiner training, as it appears to be sporadic, with newly hired staff in case study B receiving no training (5.3.4.1). Likewise, case study A data indicate that most OSCE examiners receive no formal training (4.3.4.1). According to the literature, examiner training is required for administering valid and reliable OSCEs; examiners must understand the purpose of the OSCE, the design of the marking scheme, and acceptable and unacceptable examiner behaviour (Pell et al., 2008; Reid et al., 2016; Fuller et al., 2017; Schleicher et al., 2017b; Schüttpelz-Brauns et al., 2019; Tavakol et al., 2023). The findings of this study point to a possible reason for the lack of faculty training on the OSCE: a scarcity of assessment experts. For instance, in school A, interview2-AL stated: *'I will try to bring experts and conduct courses by national and international trainers on how to perform the OSCE examination for all of our staff.'* (4.3.4.1). This demonstrates how the lack of assessment experts at these schools would negatively impact the implementation of the OSCE, including faculty training.

Using trained SPs is also necessary for valid, reliable, and authentic OSCEs; however, 'professional' SPs are preferred (Collins and Harden, 1998; Cleland et al., 2009). SPs can assist in resolving the problem of the lack of real patients in these medical schools, but neither school provide adequate training for SPs (4.3.4.1 and 5.3.4.1). For example, in some of the school A OSCEs, no training is provided for the SPs; in others, the OSCE designers provide only one hour of training before the exam begins (4.3.4.1). According to Furman (2008), MacLean et al. (2018), and Talwalkar et al. (2019) recommendations for training the SPs, these schools practise are considered insufficient because they do

not allow enough time for the SPs to practise their role and remember the scenario details, particularly for SPs who do not speak the instructor's language fluently (4.3.3.1 and 5.3.3.1). This indicates that the SPs lack professional training, which will undermine the OSCE's validity and reliability.

In terms of resources, both case studies exhibit a similar pattern: adequate educational resources but insufficient logistical resources. Participants indicated that the curriculum, teaching approach, and educational activities support using the OSCE (Hodges, 2003b). For instance, both medical schools show good examples of reviewing the students' skills before the exams (4.3.4.2 and 5.3.4.2). However, the educational value can be enhanced further by providing individual feedback (discussed above 6.2.2.3) and introducing formative OSCEs (Swanson and van der Vleuten, 2013; Pugh et al., 2018), which are both absent from both schools. Some interviewees identified that formative OSCE could also orient students regarding this assessment instrument and how to manage their time during the station (4.3.4.2).

The logistical resources, including funding, location, and equipment, are reported to be insufficient to conduct a valid and reliable OSCE (4.3.4.2 and 5.3.4.2). Although some authors deemed the OSCE resource intensive (Barman, 2005; Turner and Dankoski, 2008), Patricio et al. (2013) and Harden (2015) argued that medical education is costly, so it is important to consider the cost-benefit ratio of this instrument because, as discussed in Chapter 2, its benefits make it worthwhile. The findings of this research are consistent with the literature (Poenaru et al., 1997; Goh et al., 2022); a lack of funds and an appropriate place prevents OSCE organisers from introducing more stations to ensure acceptable OSCE reliability (4.3.4.2 and 5.3.4.2). However, approaches such as multi-site (Wilkinson et al., 2000), multi-circuit (Harden et al., 2015), multi-day, and sequential OSCE (Homer et al., 2018) have been implemented to improve OSCE feasibility. Therefore, the schools may consider administering the OSCE over three days, i.e., six stations per day, for a total of eighteen stations, but with the same resources. Nevertheless, using this approach must be accompanied by appropriate design management of student sequestering and post-hoc analysis to ensure that standards are maintained across circuits (Pell et al., 2010; Harden et al., 2015). Some participants mentioned the lack of a teaching hospital (4.3.4.2), which, if available, would compensate for some shortages in other resources by providing a sufficient place, real patients, and examiners, thereby making the OSCE more authentic and feasible. These resource constraints will inevitably force OSCE exam developers to implement low-quality OSCEs.

6.3 Opportunities and challenges to the OSCE implementation

The analysis of the findings reveals several opportunities and challenges to the OSCE’s implementation in Saudi medical schools. Despite some overlap with the previous section, the purpose of this section is to bring the discussion together to answer the

second research question, “What are the challenges and opportunities of adopting the OSCE in Saudi medical schools?”. Below, I summarise the two case studies' most significant opportunities for improvement and challenges to overcome (Table 22). The table demonstrates that although both medical schools share comparable opportunities and challenges, to a degree, school B has more opportunities than school A, which faces more challenges.

Table 22. Key opportunities and challenges of adopting the OSCE in Saudi medical schools.

	School A	School B
Opportunities	These medical schools are relatively new, so they are unencumbered by history, allowing them to perhaps improve the OSCE and implement changes more easily than older institutions.	
	The curriculum, instructional approaches, and clinical skills teaching sessions are compatible with the OSCE.	
	The desire for improvement expressed by all participants would give the introduction of change a better chance of success than if those staff members were unwilling to improve.	
	These medical schools have already collaborated with some hospitals and medical schools, so they have the potential to build on that to implement more feasible and sustainable OSCEs.	
	The majority of the staff are junior home faculty with good knowledge about the OSCE and have the opportunity to learn from the faculty development sessions.	The staff have diverse backgrounds with good experience about the OSCE, which is an opportunity to utilise their expertise to embrace creativity and innovation.
	It has the potential to obtain accreditation from accreditation agencies.	It is a private school that is more agile to change because it has more autonomy.
Challenges	There is no pre- or post-OSCE analysis to ensure the quality.	There is pre- and post-OSCE qualitative analysis to ensure the quality.
	There is no effective internal communication between faculty members and departments.	There is effective communication among the faculty members and departments.
	The lack of human resources; includes assessment experts, OSCE writers, OSCE examiners, and professional SPs. This jeopardises the OSCE's validity and reliability.	
	The financial constraint prevents the schools from securing sufficient workforce, space, and equipment to operate enough stations.	
	The absence of a university educational hospital limits access to hospital facilities—for example, difficulties in administering the OSCE in the hospital's authentic settings and obtaining real patients.	
	There is no proper psychometric analysis to ensure the quality, determine the pass/fail decisions, and identify the source of error in the OSCE implementation.	
	The need to provide individualised feedback for each student at each station to maximise the OSCE's educational impact.	
	The assessment documents lack details that guide and align OSCE implementation with the best OSCE practices in the literature.	

6.4 Synthesis of the overall discussion

6.4.1 Overarching theme: A series of dilemmas and compromises

The clustering concept is that both medical schools encounter a series of dilemmas and compromises when making any OSCE implementation decisions. The discussion indicates that the faculties made compromises at all levels of OSCE implementation without necessarily recognising them. The documentary evidence implied that both schools' written documents emphasise the basic criteria of good assessment; however, when the OSCE is implemented, these criteria are not enacted. For instance, the documents emphasise the importance of having reliable and valid assessments, but the findings of this study indicate that the reliability and validity of their OSCEs are inadequate. This illustrates the tension between theory and practice, which ultimately has resulted in suboptimal OSCEs in some respects. The preceding discussion demonstrates that the four themes generated can function as components that must be carefully evaluated to implement a successful OSCE. They provide a more holistic view than a linear guideline for medical schools to consider when implementing the OSCE. However, the overall discussion show that the current OSCE implementation in Saudi medical schools is variable and often suboptimal, necessitating a change. The following table summarises the implementation issues that have been identified from both case studies (Table 23).

Table 23. A combined list of the OSCE implementation issues that have been identified from both case studies.

1	Absence of clearly defined statements outlining the purpose and intended use of the OSCE, along with uncertainty regarding the competencies that ought to be assessed.
2	Inadequate alignment of the OSCE with the school's assessment system.
3	Lack of OSCE guidelines that follow evidence-based practices to guide the OSCE implementation.
4	Sole reliance on the OSCE as the summative performance assessment tool, with no other alternatives.
5	Insufficient blueprinting for the OSCE.
6	Inclusion of non-clinical/performance skills in certain OSCE stations.
7	Small number of stations in the OSCE.
8	Short preparation time.
9	Deficiency in assessment personnel, encompassing assessment experts, examiners, station writers, psychometricians, and professional SPs.
10	Insufficient training and briefing for stakeholders, including examiners, station writers, SPs, and students.
11	Instances of unwanted behaviour by examiners, such as interfering with students during task performance and posing questions not outlined in the marking scheme.
12	Absence of quality assurance measures, such as psychometric analysis and qualitative evaluation.
13	The standard setting is fixed
14	Lack of individualised feedback for students regarding their performance in the OSCE.
15	Language barriers in some stations affecting communication between students and SPs.
16	Absence of formative OSCEs.
17	Shortage of logistical support, including inadequate space, equipment, and financial backing for the OSCE.
18	Insufficient collaboration with other institutions, including hospitals, local medical schools, and international medical schools.

6.4.2 Public versus private medical school

There is a widespread belief that public medical schools have traditionally been well-established and state-controlled, which was thought to be sufficient to ensure high-quality assessments (Joy et al., 2007). However, the recent expansion of newly established public and private medical schools in SA raises concerns regarding the quality of medical schools' outcomes (Bajammal et al., 2008; Abu-Zaid et al., 2020). The findings of this thesis argue that the OSCE, as an example of a complex high-stakes examination, has been implemented in public and private medical schools but not always at the level of what the assessment literature would define as the gold standard. This is not uncommon; the literature indicates that OSCE is often implemented sub-optimally in different parts of the world (Bearman et al., 2017; Bearman et al., 2020).

While the discussion suggests that case study B has a more defensible OSCE implementation, both case studies have a range of challenges and shortcomings in their OSCE practice. However, both case studies have several opportunities for improvement, with a clear desire from the faculty members. Case study B's arguably better OSCE implementation can be partially attributed to the medical school's culture as a private medical school. Private schools typically emphasise quality assurance and strive to maintain a positive reputation to enrol students and maintain the budget (Altbach and Levy, 2005). This suggests that the private medical school (case study B) has better implemented the OSCE compared to the public medical school (case study A). This finding highlight that more research is needed to determine whether this situation, in which a private school outperforms a public medical school in OSCE implementation, is context-specific or if it can be broadened to other private medical schools. Such findings have the potential to significantly inform the Saudi Vision 2030's privatisation of some higher education institutions (Vision2030, 2023).

6.4.3 Key messages

The analysis and discussion proved that the OSCE implementation is challenging at both medical schools. This study reveals a series of limitations that need to be addressed by medical schools to enable the implementation of high-quality OSCEs. Firstly, the senior administration of medical schools needs to enhance the institutional and assessment culture and make it a greater priority. They should be aware of the ramifications of low-quality OSCE on a school's reputation, outcomes, and the public and students' safety. Consequently, they can value the importance of having a high-quality OSCE and invest in its improvement. The senior administration can establish a supportive institutional culture by embracing the staff through effective communication, listening to their feedback, empowering them to implement the necessary changes, and collaborating with other institutions (Simper et al., 2022).

Secondly, there is a clear need to improve the quality assurance measures used to evaluate the OSCE before, during, and after implementation. These measures include conducting a thorough review of the OSCE stations prior to implementation, inviting external examiners, and performing post hoc psychometric analysis, must address the current lack of clarity regarding the competencies that should be tested. Achieving a good OSCE design and quality assurance system requires clear and detailed assessment guidelines developed in accordance with OSCE literature that adheres to evidence-based practises.

Thirdly, the lack of OSCE personnel and their training is evident; no successful OSCE can be introduced without them. Those people include assessment experts, OSCE designers, writers, examiners, patients, and SPs. Each of them has a vital role in the OSCE's implementation; thus, they must receive adequate training and briefing. Therefore, medical schools must invest in allocating enough of them through direct recruitment or collaboration with other institutions.

Lastly, logistical and educational resources require further development. The shortage of funds, appropriate places, and adequate equipment forced the faculties to make compromises during the OSCE design, resulting in suboptimal OSCEs. Therefore, the medical schools' administrations have to secure a sufficient amount of them to enable the OSCE personnel to conduct successful OSCEs. Regarding educational resources, the two main shortcomings are the lack of individual feedback to students and the absence of formative assessment. These two elements are crucial to the OSCE's educational impact; hence, OSCE and assessment programme designers need to address these issues in order to improve the OSCE's educational value.

In summary, this thesis argues that implementing a high-quality OSCE is a multifactorial process in which all factors must be appropriately considered to achieve successful implementation. Improving OSCE quality is not only about conducting more stations or providing more examiners and logistics; while all are essential, enacting them effectively in accordance with evidence-based practices requires a supportive institutional culture. Thus, without addressing these limitations, it is unlikely that these medical schools will be able to implement OSCEs of high quality. This indicates that the awareness and commitment of institutional leaders are vital for establishing a supportive institutional culture to successfully implement a high-quality OSCE (Fuller et al., 2015). Therefore, the existing assessment theories would not be effective in this case, highlighting the need for institutional change to improve OSCE implementation, as Chapter 7 discusses.

6.5 Summary

This chapter presented a discussion and synthesis of the findings from the two case studies to answer the first research question of how the OSCE is being implemented in

Saudi medical schools. It also answers the second research question by summarising the opportunities and challenges for Saudi medical schools to implement a high-quality OSCE. In the next chapter, I will develop this discussion by presenting options for changing the OSCE implementation derived from the institutional change literature.

Chapter 7 Changing OSCE Implementation

7.1 Introduction

This chapter is an extension of the discussion chapter. It presents a theoretical perspective on how the medical schools examined in this thesis can introduce changes to their OSCE implementation to address challenges and capitalise on opportunities. I explore some change theories and then emphasise those most applicable to this thesis's context. Then, I highlight the distinctions between introducing change in public versus private medical schools and the complexity of implementing the change. I also stress the need for sustainable changes that are sensitive to the medical school context to ensure the implementation of successful OSCE.

7.2 Bringing change to the OSCE implementation

The preceding discussion demonstrates that the current OSCE implementation is, to a degree, unsatisfactory and still fails to adhere to international recommendations, so faculty members express a desire for improvement. Evidence indicates there is a strong desire for excellence, as evidenced by an eagerness to obtain national and international accreditation (4.3.1 and 5.3.1). At the same time, there is a dynamism and desire for improvement in the government, which is supported by the Saudi Vision 2030, which aims for high-quality medical education (MOE, 2022). This necessitates bringing about institutional change by utilising the limited resources, diverse contexts, and infrastructures of schools to operate affordable and high-quality, authentic OSCEs.

Institutional change is often difficult and time-consuming (Bank et al., 2017). Utilising change theory would assist medical schools in capitalising on opportunities and overcoming challenges by implementing practical steps to improve the OSCE implementation. In the next section, I will consider various change theories and explore which could be appropriate to the two medical schools' contexts.

7.2.1 Overview of change theories

I explore the change theories offered in the institutional change literature, where there are some popular change models, such as the McKinsey 7s model (Peters and Waterman, 1982) and Greiner's Growth Model (1972), which are both focused primarily on companies and businesses. The Kübler-Ross Change Curve (1969) is another well-known change model, yet it analyses change from a psychological perspective and describes people's emotions throughout the change. Moreover, the Bridges Transition Model (1995) consists of three stages that are akin to Lewin's change theory (described below), but it is rarely used in educational settings.

However, change theories such as Lewin's (1947) and Kotter's (Kotter, 1995; Kotter, 2012) are well-established in institutional change literature (Snabe, 2007; Mahmood,

2018). Both have been successfully utilised in various HPE settings (Manchester et al., 2014; van Schaik et al., 2019; Haas et al., 2020; Wijk et al., 2021). Lewin's model has been described as the core of comprehending organisational change (Burke, 2002). Kotter's change model is a more detailed approach to executing the change in higher education (Kang et al., 2022). Consequently, I deemed these theories appropriate for this research context. I aligned the two change theories because change is complex, and a single theory may not be sufficient to explain and implement the desired change. Lewin's model has three key stages: 'unfreeze, movement, and refreeze'. This aligns to Kotter's eight-step change model; Figure 16 depicts a summary of these stages.

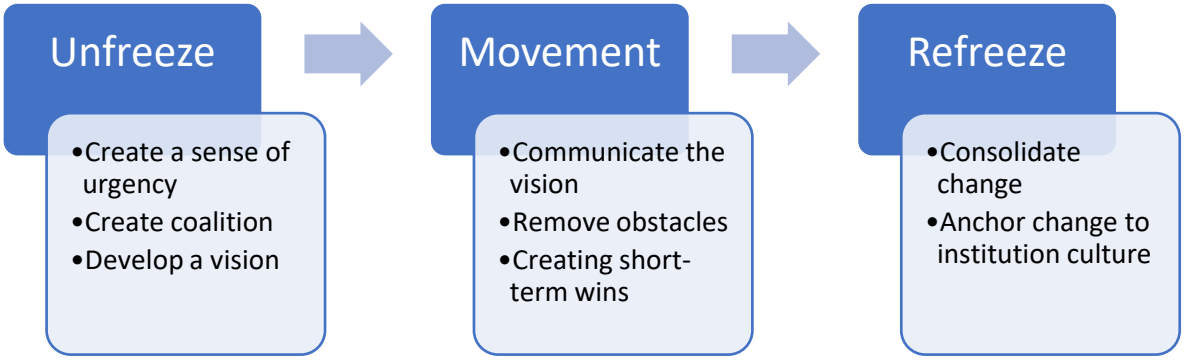


Figure 16. Stages of change. Adopted and modified from Lewin (1947) and Kotter (1995, 2012) .

7.2.1.1 Phases of change

The first phase is the unfreeze stage, which involves processes of questioning the current situation to create a sense of urgency (i.e., the right time), creating a community of people with shared needs (i.e., the right people), and developing a vision (i.e., the right plan) (Kotter, 2012; Hussain et al., 2016). These three characteristics appear to be shared by the majority of the research participants in this study. This can be reinforced by demonstrating to the faculty that change can improve their graduates and raise the school's profile. Data indicates a sense of dissatisfaction with current OSCE practices; consequently, people feel the need for administrative and staff-level changes. Furthermore, there is a desire for improvement, as many findings indicate that participants have a vision and practical ideas that need to be enabled, indicating that the faculty are committed to change. For example, an assessment specialist discussed the concept of establishing an examination and assessment centre that can be established by multiple medical schools and focuses on enhancing the OSCE implementation. This suggestion, along with similar ideas outlined below (7.2.3.2), may bring medical schools together to exchange best practises and enhance each other's OSCE. However, it is

crucial to communicate with clarity, frequency, and transparency to avoid misunderstandings and ensure a shared understanding among all parties involved. This unfreeze stage is arguably the most important phase, as it requires effective communication to convey the vision to the wider community of stakeholders.

The subsequent phase of change is the movement stage, during which the actual change processes occur. This is the time to put the change into action by taking steps to communicate and activate the vision, remove obstacles, and create short-term wins (Kotter, 2012). This is a critical stage in interpreting the vision into reality and dealing with deeply embedded issues within the culture of medical schools. For instance, the faculty development workshops that convey the change strategy will assist in communicating the vision (Kang et al., 2022). Although some data show that some faculty members are partially satisfied with their OSCE, this is only because they believe it is the best they can do, given a lack of human and logistical resources. Across the entire dataset, the lack of funds, examiners, and SPs was the most prominent issue. Therefore, medical schools could not implement better OSCEs without addressing these obstacles. It would be possible to alleviate these areas' shortages by reinforcing current efforts and expanding collaboration with hospitals and other medical schools. Medical school administrations must provide continuous administrative support, invest in human capital, and develop sources of income to maintain and support sustainable change (Kotter, 2012). Kotter values short-term achievements; thus, using key performance indicators can generate a feeling of accomplishment and support long-term success (Baloh et al., 2018). For example, findings show that OSCE stations range from two to six, so achieving more than fifteen OSCE stations in one of the end-of-year OSCEs can be used as a metric to celebrate an immediate improvement. Additionally, the data show that no psychometric analysis has been conducted, so another metric could be to start conducting an appropriate psychometric analysis. This would aid the faculty in prioritising some of these quick achievements, allowing the staff to observe the OSCE's change. This stage is filled with practical challenges that necessitate effective communication with faculty and the celebration of short-term wins in order to maintain momentum.

The final stage is to refreeze to stabilise and sustain the recent implementations. This stage entails consolidating successes and anchoring change in the school and assessment culture (Kotter, 2012). The findings indicate a lack of faculty development sessions; therefore, senior and newly joined faculty must receive training in their new roles to provide stability for the change, as they are the most critical asset for sustaining the change. Additionally, the analysis of the documents reveals a lack of OSCE-specific policies and procedures; therefore, all systems, policies, and regulations must be revised to reflect the new reality (Odiaga et al., 2021). This stage is crucial for solidifying the change and influencing the institution and assessment culture (Ndoye and Parker, 2010). While this phase concludes the big change, no change is error-free; therefore, evaluation

channels should always be available to identify and address shortcomings (Rhydderch et al., 2004). When the staff witness changes resulting from their evaluation, this reinforces the change and encourages additional evaluation, thereby sustaining the improvement cycle.

7.2.2 Change in public versus private medical school

This period in Saudi medical education history is characterised by a dynamic transformation in higher and health professions education that runs parallel to the government's vision (Aljohani, 2020). The two most important aspects of Saudi Vision 2030 in this regard are the graduation of competent healthcare professionals and the privatisation of some public universities (Alnufaiee, 2019). This aims to empower universities to develop and achieve financial balance and sustainability (Vision2030, 2023). These issues would motivate medical administrations to make changes in order to meet the national development strategy. Therefore, the number of private medical schools in Saudi Arabia is increasing (Ayuob et al., 2015).

While the improvement of OSCE implementation is possible in both types of medical schools, private medical schools may have more space to accommodate agile change (Mohamed, 2003). In contrast, public schools may encounter a hierarchical decision-making process and limited resources to make changes (Farghali, 2013). This was highlighted by the academic leader in the private school (B) (5.3.4.1), who suggested he would recruit assessment experts, indicating that such a decision can be made quickly when the benefits are evident. Hamron (2018) and Alghamdi (2019), in their research on the privatisation of higher education in Saudi Arabia, conclude that greater financial and administrative autonomy enables private schools to collaborate with national and international institutions more easily. Althubaiti (2020) studied the new Saudi government's vision to privatise higher education and indicated that it is part of Saudi economic reform for the sector. While several studies suggest that the privatisation of higher education may have a negative impact on its quality (McCowan, 2004), Althubaiti (2020) argues that privatisation may contribute to economic development, meet the requirements of the labour market, and enhance the quality of educational services. Furthermore, the findings indicate that the private medical school has better internal communication and external collaboration with other institutions. It also has an OSCE-specific committee that monitors its quality and senior staff who have received OSCE training at various institutions. These factors may make this private medical school more prepared to implement the change for their OSCE implementation.

Althubaiti (2020) expressed concern that private medical schools might prioritise profits over educational quality. However, the findings of this study indicate that the private medical school in case study B have more measures to control the quality of their OSCE and better internal communication. Nonetheless, addressing the concerns of

stakeholders can be accomplished on two levels: firstly, by increasing the quality criteria for the accreditation of medical schools (Alrebish, 2014; Abrol, 2016). Secondly, explore the impact of an OSCE national licencing exam after graduation in addition to the current written licencing exam to ensure the clinical competence of new graduates (Abu-Zaid et al., 2020). However, there are different models to ensure the clinical competency of new graduates (Price et al., 2018); one model is the UK model, which is a national set of clinical and professional skills assessment (CPSA) standards for medical schools (GMC, 2023a); another model is the national clinical skills exam such as that in Switzerland (Bonvin et al., 2019; Huwendiek et al., 2020); and a third model is that of the United States (USMLE, 2023) and Canada (MCC, 2023), as they formerly had a unified national OSCE exam that is currently discontinued, although the specific reasons for this decision remain unclear. This overview reflects the debate in the literature regarding the best approach to ensuring medical graduate clinical competencies (Swanson and Roberts, 2016), which calls for a Saudi national study to evaluate the Saudi context, examine the experiences of other countries, and provide recommendations for the Saudi context.

Enhancing the quality of private medical schools can be expensive, which can result in an increase in student fees. This will make public medical schools more appealing to prospective students, given their long history of public trust (Alrubai, 2011; Alzahrani, 2019). As a result, a decline in the number of applicants to private schools may prompt these institutions to lower admissions standards in order to maintain financial stability. Therefore, to assure the public, the above recommendations regarding accreditation agencies and licencing exams are worth considering.

The findings indicate that there is a need for substantial change at the public school (A) due to a sense of demotivation among some staff members and poor internal communication (4.3.1). The public school has more bureaucratic processes to change things. Thus, the school needs to focus on accelerating the change approval process, as staff naturally want to see change quicker. Change should not be limited to the introduction of specific changes, such as a new approach to standard setting, but should instead focus on achieving a more agile approach to supporting change processes in general. On the other hand, the more mature OSCE delivery at the private medical school may suggest that their system effectively manages some of the OSCE's complexity. This presents an opportunity for them to build upon their system to improve their assessment quality. The dilemma and tensions in implementation are noticeable in both private and public types of medical schools. Therefore, change is not simple, as it involves more than just improving technical metrics and must consider educational, economic, and contextual factors along the way.

7.2.3 Complexity of change

H.L. Mencken once said: “There is a solution to every problem: simple, quick, and wrong.” (Nelson, 1976). In the literature on medical education, complexity theory has been utilised to illustrate the nonlinear interactions between multiple variables when designing and implementing educational programmes (Jorm and Roberts, 2018; Cristancho et al., 2019). Furthermore, medical education is a multidisciplinary field with origins in medicine, education, administration, and other fields. For example, implementing OSCE is a complex task that is influenced by institutional culture, workforce availability, staff experience, policy decisions, infrastructure, and funds. Therefore, OSCE implementation change is not merely an academic change but a combination of institutional, contextual, and interprofessional change. Additionally, many stakeholders may be resistant to change, making its implementation more complicated (Bajammal et al., 2008; Harrison et al., 2017). Change implementation is usually complex and challenging; therefore, the change leader must be sensitive to the organisational culture, be flexible, and strive for sustainable change.

7.2.3.1 Sensitivity to the context

In the literature on institutional change, sensitivity to institutional context has been emphasised (Buchanan et al., 2005; Beach, 2016). While some developed countries may be able to adapt to the structure and situation of private medical schools, the Saudi context is different. All old schools are public, but in recent years the number of private schools has increased, and many public schools have begun to be privatised (MOE, 2022). This would have implications for the OSCE as an example of a costly and complex performance assessment. According to the data, this presented numerous financial and logistical challenges. Consequently, utilising the experiences of international medical school partners should only be considered in relation to the context of local settings. While international partnerships can provide valuable insights and standards of practice, it is essential to consider local cultural, social, economic, and assessment system factors to ensure their effective implementation and applicability. Because academic institutions have different staff expertise, funding, vision, culture, and logistics, all of these factors will play a crucial role in implementing the change (Buchanan et al., 2005). The data indicate that in contrast to case study A, case study B's collaboration with international medical schools resulted in a better OSCE implementation. However, some OSCE scenarios must be revised to reflect the local context (5.3.1.1). Furthermore, case study B OSCEs have certain shortcomings that render their implementation inferior to the international school, such as a small number of stations and the use of a fixed cut-off score.

Despite the data indicating a desire for improvement, there is still a sense across both schools of underestimating the OSCE’s complexity and difficulty in implementing change

in the practice. Participants desire to increase the number of stations, provide training for stakeholders, increase the number of examiners, and other interventions that are valuable and essential for the success of the OSCE. However, introducing substantial changes to each level of OSCE implementation necessitates establishing a solid foundation first. As all factors are interconnected, separate modifications may not be sufficient to achieve a high-quality OSCE. For example, case study B wanted to increase the number of stations from six to twenty-six, so they introduced unobserved OSCE stations, which, as previously discussed (5.3.3.2), do not constitute an OSCE as commonly understood. However, they felt compelled to do so because of a shortage of station writers, examiners, and logistics. Therefore, a leader with medical education assessment knowledge, vision, authority, and a holistic approach to change is vital for successfully implementing the change. This is not the sole responsibility of a single person, such as the chair of the OSCE Committee, but it needs to involve a team with the passion and expertise to make positive contributions (Sanyal and Hisam, 2018). It requires the team to return to the barriers and levers, considering the system as a whole and how the OSCE would fit into it (Khan et al., 2013a). An example of that could be looking for a comprehensive approach to the assessment system, so instead of one exam for each course, an integrated, cross-modular OSCE may be more appropriate.

7.2.3.2 Sustainable change

From the complexity theory perspective, Mason (2009) stresses the importance of sustainable change. Considering sustainability when planning for change is critical to avoid relapsing into old problems (Olafsen et al., 2021). Some participants, for example, stated that they had previously received OSCE training that had been discontinued (4.3.4.1 and 5.3.4.1). My inquiry into this matter revealed that the partnership contract with the international experts who provide the training was modified, and the training was discontinued. While it is advantageous to utilise external expertise, this temporary situation requires the administration to invest in developing local expertise. Local experts would contribute to the training's sustainability and aid the OSCE's implementation due to their understanding of sociocultural nuances.

Change that is more effective and long-lasting should be built on the local institutions' capacities rather than being imposed on people. Consequently, there is a role for cooperation between regionally competitive medical schools, i.e. co-opetition (Hidayah, 2018). Co-opetition can empower institutions to collaborate and capitalise on each school's comparative advantages in order to create feasible, sustainable, and high-quality OSCEs (Muijs and Rumyantseva, 2014). The Australasian Collaboration for Clinical Assessment in Medicine (ACCLAiM) project acknowledges that establishing a high-quality OSCE can be difficult and costly. So they came together to share the best practices for creating the OSCE stations, such as writing the stations, training examiners, analysing the results, and benchmarking (ACCLAiM, 2023). Another option is an online

forum that brings together individuals with a special interest in the OSCE to share expertise and enhance one another's practises, akin to the psychometrics special interest group organised by the Association for the Study of Medical Education (ASME, 2023). Likewise, the Saudi Medical Colleges Deans' Committee (SMCDC) has an opportunity for collaboration to facilitate high-quality yet affordable OSCEs. Such projects have a high chance of success because they adhere to the Saudi Vision 2030 principle of collaboration between the public and private sectors (Vision2030, 2023).

7.2.3.3 Flexible change

Flexibility in implementing the change is essential to a successful change (Dunford et al., 2013). For example, increasing the number of stations per exam from two or six to twelve would necessitate a significant resource that would not be readily available. Therefore, the change takes time to evolve and is subject to unpredictable factors with no straightforward recipe that will guarantee success (Styhre, 2002). However, with a clear vision, the support of the administration, and the staff's commitment, a desired change can be accomplished (Shanley, 2007). The implementation of new assessment standards is difficult but necessary. Therefore, the design of the change should be sensitive to the organisation's culture, sustainable, and flexible.

In addition, evaluation and dissemination of the implemented change are crucial for ensuring continual development and dissemination of best practices throughout the medical education community (Graham et al., 2003). Evaluation is essential to evaluating the change's effectiveness, enabling schools to identify areas for improvement and make informed decisions about what they should do to improve further (Harden et al., 2015). Dissemination is also important to share the experience and the process of change with other stakeholders and institutions about evidence-based practices (Ravinetto and Singh, 2022). This can be accomplished through publications, conferences, workshops, and collaboration with other institutions.

7.3 Summary

This chapter offered a theoretical perspective to introduce institutional change to support the implementation of a high-quality OSCE. It presents change theories that will be useful in guiding the desired change at the level of educational institutions. It highlights the differences between public and private medical schools when introducing change, the complexity of change, and the need for sustainable, flexible, and context-specific change. In the following chapter, I will present the thesis's conclusions, the contributions of this thesis to the literature, recommendations for different stakeholders, strengths and limitations, and conclude with a reflective account of this research.

Chapter 8 Conclusions, Recommendations, and Reflections

8.1 Introduction

This chapter presents the final sections of the thesis, including the conclusions, contributions to the field, and recommendations. I present the recommendations in three sections: medical school recommendations, national and international medical education recommendations, and recommendations for future research. Then, I discuss the strengths and limitations of this thesis. I conclude my thesis with a reflective account of this project.

8.2 Conclusions and a review of the thesis

The literature generally highlights that the OSCE is one of the most valid and reliable performance assessment instruments when appropriately implemented (Abdulghani et al., 2015; Harden et al., 2015; Boursicot et al., 2020). Therefore, good implementation is the bedrock of a successful OSCE (Khan et al., 2013b). This thesis is the first explorative and interpretive study to explore how the OSCE is being implemented and identify the challenges and opportunities its adoption offers in Saudi medical schools. Whilst the literature review (Chapter 2) demarcates the criteria of good performance assessment, there is a paucity of research on how medical schools in new contexts, such as Saudi medical education, implement and ensure the OSCE's quality.

The data was gathered from two case studies conducted at two newly established medical schools, one state-funded and the other privately funded. This study's multi-method data collection and analysis approaches provide valuable insight into the complexity and disparity in the quality of OSCE implementation in the Saudi context. The inductive RTA analysis generated four cross-cutting themes that contribute to this situation: institutional and assessment culture, faculty expertise and practices, OSCE quality and design, and resources and infrastructure setup. However, the overarching theme across the two case studies themes and sub-themes is that there are a series of dilemmas and compromises at each stage of OSCE implementation.

Even though the assessment document analysis at both medical schools shows high congruence with the criteria of good assessment, it appears that these criteria are usually not reflected in practice. According to the findings, many staff members are aware of the essential requirements for effective OSCE; nonetheless, several reasons appear to hinder them from adhering to the OSCE standards described in the literature. For instance, it is unclear what competencies the OSCE is supposed to assess and what its intended purpose is. The data interpretation revealed that contextual aspects such as the funding source of the school (private or public), accreditation status, faculty experience, and availability of resources are linked with the quality of the OSCE implementation.

The overall aim of the research is to provide insight into possible interventions that enable medical schools to implement high-quality OSCEs. Data shows that most faculty members are eager to improve their OSCE practices; however, the institutional culture, workforce, infrastructure, and financial resources are not supportive of implementing OSCE that meets high-stakes OSCE quality standards. Therefore, I propose introducing changes to the medical school to improve OSCE implementation and align with the Saudi government's 2030 vision for enhancing higher education.

Based on this work, I recommend introducing change to improve the OSCE implementation using well-established change theories that are applicable and effective in educational settings. The proposed changes at the medical school level may include redesigning the assessment system, investing in human capital and infrastructure, and employing quality assurance measures. More change can be introduced at the national and regulatory bodies level, including developing a national roadmap to facilitate performance assessments, encouraging coopetition between national medical schools, partnering with hospitals and international medical schools, and forming a team of national assessment experts to support OSCE implementation. However, the discussion underlined that introducing change is a complex and time-consuming process; thus, change leaders should be flexible in implementing the change and seek sustainable transformation rather than short-term success.

While both medical schools examined in this study can implement a successful change to their OSCE practice, the private school may have some advantages when bringing about the change. Findings suggest that private schools may have a better institutional and assessment culture, such as being more agile in introducing change due to the flexibility of their administrative structure, having better internal communication, having more senior staff, and being eager on quality assurance to maintain the institution's reputation to maintain the revenue flow from student admissions.

This thesis contributes to the body of knowledge by offering an original, in-depth understanding of the OSCE implementation in Saudi medical schools. It recommends introducing change to exploit opportunities and overcome challenges to improve the OSCE implementation. Moreover, this project provides a novel perspective by highlighting the link between the source of school funding and the OSCE implementation in the two medical schools examined in this thesis. Another contribution is using an innovative combination of methods, and analysis approaches to answer the research questions effectively. Additionally, borrowing change theories from other disciplines revealed that medical education could benefit from interdisciplinary research to improve discipline-specific issues like OSCE implementation.

8.3 Contributions of this work

This thesis has made several contributions to literature knowledge and theory, as demonstrated in the following four sections.

8.3.1 In-depth exploration

Firstly, this thesis is the first in-depth research exploring the OSCE implementation in Saudi medical schools using a large-scale multi-method qualitative project. This established an original understanding of how the OSCE is implemented in this context and offered insight into the complexity of implementing a performance assessment. The research identifies a series of challenges to overcome and opportunities to exploit to improve the OSCE implementation. The findings reveal a sense of tension and a series of dilemmas at each stage of OSCE implementation. This suggests the need for an institutional-level change to improve OSCE practice. The proposed change can occur in different areas, such as regulations, assessment programmes, human resources, and school infrastructure investment. Furthermore, the findings from the implementation of OSCEs in the investigated medical schools can provide the groundwork for the proposed national OSCE exam for medical graduates (Abu-Zaid et al., 2020). Another alternative is to introduce a national set of standards similar to the GMC's for the CPSA that ensure the quality of the OSCE (GMC, 2023a).

8.3.2 Novel perspective

Secondly, this study advances knowledge by offering a novel perspective by examining the impact of institutional culture, faculty, and logistics on OSCE quality and design at two medical schools (private and public). Whilst both have similar challenges, the private medical school (case study B) appears to have some factors like effective internal communication, which contributes to better OSCE practices. Furthermore, this private school appears to be more adaptable to change because their decision-making process is less bureaucratic as a result of their financial independence.

Therefore, this thesis contributes to the body of knowledge by providing an original understanding to view the OSCE implementation through educational, economical, organizational culture, and health quality lenses. Educationally, controlling the OSCE's quality would have educational ramifications for students' pass and progression, ultimately, their graduation or failure and remediation. Economically, this study establishes a connection between some OSCE challenges and funding issues by demonstrating the logistical and recruitment costs associated with OSCE use. The organizational culture implications are numerous, including administrative and communication aspects. However, patient safety is paramount because it is the ultimate aim of this process, for which all resources must be exploited to tackle all challenges to achieve this goal (Tamblyn, 1999). Budget and poor quality assessment are not only a

threat to patient safety but also to learner safety, as it would provide learners with a false sense of reassurance that they are competent to practise (Smirnova et al., 2019).

8.3.3 Unique methodological combination

Thirdly, I believe that one of the most significant contributions of this study is the use of in-depth case studies (Yazan, 2015; Yin, 2018) involving the combination of multiple methods, which provides new methodological lenses for gaining a holistic understanding of the research problem. This thesis used a novel combination of qualitative methods to investigate OSCE implementation, combining Codebook analysis for the documents with RTA for the interviews and FGs. Then, integrate the findings into the discussion to offer answers to the research questions and recommendations for medical schools. Triangulating different data sources, employing different methods, and using different analytical approaches produce a new dimension in understanding OSCE implementation. This approach demonstrates a useful application for analysing my research data, so I would recommend it to other researchers working in various contexts and with various assessment instruments. This new methodological lens demonstrates, for instance, that assessment documents may describe something that is not implemented in practice and vice versa.

8.3.4 Introducing change theory

Fourthly, this thesis contributes to the OSCE literature by introducing change theories to improve their implementation (Chapter 7). It also raises awareness of the importance and complexity of implementing a high-quality OSCE that meets the criteria of good assessment. The proposed change theories emphasise the need for sustainable, flexible, and context-sensitive changes to improve and maintain the quality of the OSCE. Furthermore, it provides empirical evidence for medical schools worldwide, particularly those in similar contexts, about the challenges they must be aware of, how to address them, and the opportunities they must seize. It highlights areas in the OSCE implementation that indicate whether their OSCE is of high quality or needs to be improved. It shows gaps in some OSCE practices that need to be addressed to introduce a defensible OSCE. Therefore, I believe that this thesis provides actionable steps for change and recommendations to enhance OSCE implementation in medical and health schools worldwide.

8.4 Recommendations and implications

In practice, this thesis proposes several recommendations involving what changes can be made to improve OSCE implementation and, more broadly, assessment and education in various medical and other health professions schools worldwide. I believe this study’s findings and recommendations will have significant implications for medical schools in SA and other countries with similar contexts.

While developing these recommendations, I considered the tension between producing a high-quality OSCE and the ever-present resource constraints. Therefore, I consider leveraging available resources judiciously to achieve implementation feasibility while offering high validity and reliability for the OSCE. Although there are some differences between the two medical schools in this thesis, the recommendations are overarching and applicable to both, and arguably the majority of this could be applied to most medical schools that share their contexts.

8.4.1 Recommendations at the medical schools' level

- **Redesign assessment programmes** and course schedules to ensure the OSCE is nested within a suite of appropriate, authentic assessment instruments supporting education. Based on Kotter's theory of change (Kotter, 1995), this recommendation could serve as the change vision, outlining the proposed change. The assessment programme should place greater emphasis on assessing all relevant skills, which need to be aligned with the intended learning outcomes and assessment instruments. The assessment programme should incorporate multiple assessment instruments, such as the Mini-CEX, DOPS, and long-case examination. The OSCE should introduce a greater number of integrated stations that can help mimic the undifferentiated practice of new doctors. The current practice of introducing an OSCE exam for each course would allow for only a limited number of skills to be blueprinted and examined. Restructuring the assessment programme should consider allowing more skills to be blueprinted and examined. This could be accomplished by designing one more authentic, integrated, and long OSCE exam per year for all courses. That will allow for better preparation, blueprinting, the introduction of more stations, the ability to secure adequate examiners and SPs, and better logistical management for the limited resource. This would result in OSCEs that are more valid, reliable, and feasible.
- **Develop assessment policies and regulations** to align with the criteria of good assessment, evidence-based practices, and up-to-date OSCE literature. This recommendation would be consistent with Kotter's theory of institutional culture change consolidation through documentation. Efforts to improve the assessment documents should concentrate on including more information about OSCE implementations. This should include, but not be limited to, clearly stating the intended purpose for the OSCE use in relation to the assessment system, listing the types of competencies that the OSCE can assess, outlining the minimum criteria to run the OSCE, deciding an appropriate method to calculate standard settings, developing a procedure for providing feedback to the examinee, and providing guidance for quality assurance measures. This would help avoid OSCEs that do not meet the basic OSCE criteria, such as exams with few stations, unobserved stations, and no individual feedback. The OSCE is a resource-intensive examination, so

administering it inappropriately would waste resources and produce indefensible results. OSCEs are intended to assess observable clinical skills, and knowledge and reasoning can be assessed in an OSCE, but only as part of an observable clinical encounter. Therefore, management should make sure these documents are disseminated and followed by all OSCE personnel.

- **Invest in human capital,** as there is a legitimate need for more OSCE staff. This can be achieved by recruiting assessment specialists, OSCE examiners, OSCE writers, and SPs. Each of these personnel is in high demand, particularly assessment experts who can inspect OSCE implementation, oversee the entire assessment system, and train faculty. A sufficient number of workers is essential for valid and reliable OSCEs, as no real improvement can be implemented without them.
- **Focus on faculty development activities,** including training sessions for OSCE designers, writers, and examiners. Training would empower the action step of Kotter's change theory because it would enable individuals to act towards the vision. Training is essential, so it should not be taken lightly or underestimated how much time and effort should be put into this endeavour. While workshops should provide theoretical background for the OSCE, it is important to include hands-on training, such as station writing and student assessment. Having qualified examiners is vital for valid and reliable scoring. Therefore, outside examiners invited to cover examiner shortages should receive comparable training to understand their role and have the appropriate expectations of the students. Furthermore, the examiners must be trained to provide each student with effective, personalised, constructive, and actionable feedback.
- **Provide proper training for OSCE participants**, including candidates, SPs, patients, and support staff. A briefing before the exam would not suffice, as proper training requires time and spaced-out instructions to ensure participants can perform as intended. Candidates need to engage with formative OSCE to get used to this format, ensure they understand the different types of stations and learn how to manage their time effectively within each station. Training SPs, patients, and support staff ensure consistency and standardisation in the OSCE delivery, enhancing the exam's validity, reliability and fairness. Moreover, because training new people for each exam can be challenging, creating a local database of SPs and patients would be more feasible.
- **Use quality assurance measures** to ensure the quality of the OSCE and inform curriculum, teaching, and future OSCE implementation. This can be accomplished by starting an OSCE evaluation programme incorporating quantitative and qualitative data. Quantitatively, post-hoc psychometric analysis can provide invaluable information about the OSCE’s quality, such as identifying the source of implementation errors. In addition, benchmarking students’ results to those of their peers from other medical schools can inform the institution about the performance

level of its faculty and students. Qualitatively, receiving feedback from various stakeholders. For instance, feedback can be received in writing or during a post-exam discussion session where examiners and examinees discuss and analyse exam items. This would provide both qualitative insights for the faculty and educational benefits for the students.

- **Provide additional administrative support and improve communication with different departments** to ensure a quality assessment. The administration's authority is critical in implementing and maintaining all recommendations. According to Kotter's theory of change, communicating the vision to all stakeholders is crucial for the success of any change. Furthermore, celebrating short-term achievements upon reaching a predetermined milestone would increase confidence, momentum, and support for the change efforts. Furthermore, they can work to increase the OSCE budget by securing adequate funds, space, and equipment. They can also facilitate the establishment of a teaching hospital. This would enhance bedside teaching and make the OSCE more authentic by introducing it in real settings with real patients.

While these recommendations are applicable to both public and private medical schools, each type of institution may need to emphasise specific recommendations. For example, the data indicate that the public school may need to improve their internal communication further and that the private school may need to invest more in developing comprehensive assessment policies and regulations.

8.4.2 Recommendations at the national and international level

These are the recommendations for regulatory bodies and other institutions involved in medical education worldwide.

- National quality assurance bodies, such as the NCAAA in SA, have the opportunity to **design a quality framework for performance assessments** to guide the OSCE and other performance assessment tools. This would allow them to collect and evaluate data from medical schools, ensure the quality of assessment implementations, and suggest possible interventions medical schools can employ to overcome challenges in OSCE implementation. For example, they can produce a clear, practical, and up-to-date OSCE guideline that provides schools with information, recommendations, and instructions for implementing high-quality OSCEs. Developing OSCE implementation guidelines does not imply a static implementation devoid of creativity and innovation; rather, it merely establishes the minimum acceptable standards for good OSCEs. This can be reinforced by regulatory body visits to evaluate implementation on the ground and promote best practices. This recommendation would serve as a foundation for evidence-based practice and a powerful change agent.
- Exam providers for medical licences, such as SCFHS in SA, may need to **evaluate the value of the national performance licencing examination**. For example,

introduce an OSCE performance exam alongside the current written exam. This would ensure medical graduates' clinical competencies for the sake of patient safety. However, there are different approaches to consider, such as introducing a common set of OSCE standards, as indicated in the preceding chapter. This effort would improve undergraduate OSCEs by establishing standards and encouraging medical schools to improve their OSCE practises in accordance with the national standards.

- The Saudi Medical Colleges Deans' Committee (SMCDC) and similar committees in different countries have the opportunity to **develop a collaboration roadmap to facilitate performance assessments**. It is recommended that partnerships be formed with regional medical schools to **establish a joint simulation and examination centre** for teaching and examining students as well as training staff. Although competition between medical schools is inevitable, mutual interest would motivate decision-makers to establish such initiatives. This can be considered a coopetition that would resolve logistical issues such as a lack of expertise, workforce, funding, space, and SP. So, instead of having a smaller version of this centre in each medical school (i.e., clinical skills labs), it would be more cost-effective for all partners to share resources and costs for a well-established advanced centre.
- **Form a coordinating council** between regional medical schools to share expertise and **form a network of Saudi medical school OSCE experts.** As they share similar contexts, they can assist each other in operating the OSCE and discuss overcoming challenges and exploiting opportunities**.** Members of this council can collaborate as a joint department to provide expertise in psychometric services across multiple schools, build a shared OSCE station bank, and promote further collaboration. This concept can be expanded to establish an assessment support team to provide medical schools with research services and evaluation for assessment programmes. Due to the paucity of assessment specialists in the region, this would greatly benefit medical education.
- **Consider partnering with hospitals and national medical schools** to work together, support and improve OSCE implementation, and establish a network-based development approach.
- **Launch a journal and an annual conference on assessment in medical education** to inform and promote advances in assessment in similar medical schools.

8.4.3 Recommendations and directions for future research

After completing this research, I believe that its findings highlight numerous future research potentials in this field. Firstly, I recommend that future research utilise the methodology of this thesis to investigate the OSCE implementation across different medical school contexts, such as medical schools in the Arabian Gulf countries. This would inform the field whether the findings identified in this dissertation are comparable

across medical schools in the region or whether they are highly context specific. There is scope for future researchers to learn about these various organisations and apply this thesis's lens to understand the effects of various organisations' structures on assessments and overall school outcomes. This can determine whether OSCE implementation disparities are attributable, at least in part, to the fund source or other contextual factors.

Secondly, this thesis recommends using certain change theories to improve OSCE implementation; however, I came across many change theories in the change literature. For example, the McKinsey 7s model (Peters and Waterman, 1982) and Greiner's Growth Model (1972) are well-known models of change in business but not in educational literature. Therefore, it would be beneficial for future research to examine what change models might be suitable for introducing the desired change in medical school assessment and whether different school structures require a different change model. This would increase our understanding of how the OSCE can be best implemented in different contexts.

Thirdly, the findings of this thesis suggest a disparity in the quality of OSCE implementation between institutions in SA. This suggests the importance of ensuring the quality of performance assessments used to graduate new medical doctors. Some may argue that this calls for evaluating the need for a national licensure OSCE exam in Saudi Arabia to ensure the competency and safety of newly-graduated physicians (Abu-Zaid et al., 2020). This requires a research team conducting a systematic review of international practices as well as a local environmental screening to determine the need for such intervention or to look for other alternatives.

Fourthly, this thesis highlights the importance of conducting an exploratory study to evaluate the wider assessment culture and assessment programmes in SA medical education. The OSCE, like any other assessment instrument, is merely one component of a larger assessment system (Eva et al., 2016; Norcini et al., 2018); thus, researchers can investigate how assessment systems are structured and implemented in Saudi medical schools. Understanding this issue in greater depth can provide crucial data regarding the quality of medical education and the outputs of medical schools.

Fifthly, the analysis indicates that this private medical school has a higher OSCE quality than the public medical school. Therefore, future research may investigate to what extent private school characteristics, such as a focus on customer service, contribute to this situation – and whether this difference is more broadly typical. Furthermore, whether the differences discovered in the two case studies are applicable to other medical schools in SA (e.g., old, well-established, and state-funded) as well as other schools' aspects beyond assessment and OSCE (e.g., quality of teaching).

Finally, according to Connell (2014), Southern theory asserts that context differences make knowledge experiences different due to the north-south global disparity. This theory challenges the inherent dominance of Western perspectives in social and academic knowledge. Therefore, it calls for including perspectives from the Global South to create a comprehensive understanding of the issues under consideration. While this research is in a Southern context but uses Northern guidance to critique the OSCE, future research could examine whether there should be more specific Global South guidance and examine the OSCE through that lens.

8.5 Strengths and limitations of this thesis

The sections that follow provide an overview of the thesis's key strengths and limitations.

8.5.1 Strengths

I identified several strengths in this thesis upon reflection. First, the research questions addressed gaps in the literature about OSCE implementation in Saudi Arabia. The literature review I presented in Chapter 2 reveals a dearth of studies on the quality of OSCE implementation in international literature, but none in the Saudi context. Therefore, this is a key strength for this thesis as it explores the uniqueness of this study context from other Western countries, distinguishing it from the vast majority of OSCE studies discussed in the literature. I believe this in-depth investigation would enable decision-makers, medical educators, and OSCE workers who share this thesis' context to rethink their practises and identify the drivers and barriers to improving the OSCE implementation.

Secondly, I adopted the constructivist qualitative paradigm in this thesis to investigate the OSCE implementation. In the field of assessment, which is dominated by quantitative research and lacks realist research methods (Hodges, 2003a; Norman and Eva, 2018), the constructivist qualitative approach is viewed as a new paradigm with the potential to bring significant advantages to Saudi Arabian medical education (Farghaly, 2018). According to my literature review, this is the first time the constructivist paradigm has been utilised to explore in detail the implementation of the OSCE in two contrasting contexts. This research can contribute to introducing this paradigm to the Saudi medical education community by demonstrating its applicability and utility to practice. Rather than focusing on the surface of an issue using the positivist paradigm, the constructivist paradigm enables the researcher to delve deep into its origins (Tavakol and Sandars, 2014; Cohen et al., 2018). I believe this paradigm is valuable for addressing similar research questions; therefore, I anticipate this thesis to promote its use in Saudi medical education research.

Thirdly, this research methodology is designed to triangulate the data in two ways. Firstly, it was collected from various sources: documents, academic leaders, assessment

specialists, OSCE organisers, OSCE writers, and OSCE examiners from two medical schools. Secondly, the data is triangulated utilising three qualitative research methods: document analysis, interviews, and FGs. Furthermore, this thesis employed the appropriate data analysis approaches for each form of data, namely Codebook analysis for the documents and RTA for the interviews and FGs. Combining these data sources, research methods, and data analysis approaches enriches the findings by looking at the OSCE implementation from different perspectives. This methodology enables the production of findings and in-depth analyses with a high degree of validity and robustness, thereby enhancing the credibility of the findings.

Fourthly, I conducted all the research procedures myself. This includes the recruitment of participants, the conduct of interviews and FGs, the processing of data, and all phases of analysis and writing. This offered me the chance to immerse myself in the data, allowing me to gain a comprehensive yet in-depth understanding of the context and OSCE implementation in Saudi Arabia. I also share the participants' culture, allowing me a greater understanding of their perspectives and linguistic expressions. The aforementioned features are valuable components of adopting the constructivist/interpretivist paradigm. Consequently, my perspective is inextricably connected to the study, as I have to engage in all research procedures. For instance, during that interaction with the participants, I observed that they were discursive on some issues but not others, which is a valuable point. I consider this a good sign for two reasons: firstly, it reflects the strong rapport I have established with them, which allows them to communicate openly and honestly. Second, when they begin to engage in discourse about an issue, it is a sign that it is a significant issue that I should investigate further. In fact, my unique context allowed a deeper, more nuanced interpretation. Much of what I have encountered has been non-verbal, and by doing all the work myself, I bring continuity to interpretation that would be lost with an assistant, for example. All of these factors enhance the utility and precision of the data.

Fifthly, this thesis took place during an unprecedented period of reform in Saudi Arabia initiated by the 2030 Vision, which brought about changes in a variety of areas, including education, medicine, and the economy. The number of medical schools in Saudi Arabia has increased rapidly from five to forty-one in the last twenty-two years, as discussed in Chapter 1. This increases competition among these medical schools, and only those with high-quality outcomes and alignment with the 2030 Vision, which this thesis supports, are likely to stand out to prospective students and the government. Given the topic's importance to the medical and educational sectors, this research will help in the development of good assessment practices in these fields. Thus, this study will be a valuable addition to medical education in Saudi Arabia.

8.5.2 Limitations

Although this project adhered to methodological rigour and addressed the research questions and objectives, it is not without limitations. Potential limitations can relate to methodological and communication issues, which I will address in turn.

8.5.2.1 Methodological issues

This thesis is an empirical study of two medical schools, but investigating more medical schools might aid in examining a broader context, especially when pointing to comparisons between public and private institutions. For instance, since both case studies are considered newly established, including an old and well-established medical school may be advantageous. However, my investigation was focused on depth and breadth as opposed to a superficial investigation, and the limited time and resources—i.e., being a single researcher constrained by PhD time constraints—restricted my ability to investigate additional case studies. Nevertheless, from a philosophical standpoint, qualitative research investigates issues in more depth than breadth (Cohen et al., 2018). Therefore, the amount of data collected falls within the range of comparable qualitative studies in the literature and is deemed adequate for the methodological paradigm employed. Importantly, the data collected enables me to answer the research questions, address gaps in the literature, and generate sufficient supporting evidence.

Similarly, the nature of the PhD, which makes me the sole researcher with a limited amount of time, influences some decisions and is the source of certain constraints, such as the inability to observe actual OSCEs due to practical restrictions. Another example is that this thesis has no quantitative component, such as statistical or psychometric analysis. In fact, the initial design of the study included both quantitative and qualitative analyses of the OSCE data collected from these case studies. However, it became quickly apparent that this would be an excessive amount of work; furthermore, subsequent findings suggested that barely any psychometric evaluation had actually taken place. As a result, it is clear that this is not a feasible design for a PhD project. Nonetheless, the quantitative component may still add some value to the overall findings, so I would like to conduct supplemental qualitative research to support this work further.

Purposive sampling is another methodological limitation that affects the case study selection process. I set certain requirements for case study selection, including being from different cities and having different funding models, to enable me to examine how the OSCE is implemented in different contexts. Although the case studies in the thesis satisfied these requirements, the selection for medical school depended in part on the accessibility to the medical school. While selecting alternative medical schools may impact the data and analysis, the findings in the contexts I have examined are clear. Since the medical schools in SA share a broader context, I believe that the methodological approach, findings, and recommendations can be applied to a number

of medical schools in the country. Furthermore, medical schools outside of SA with similar contexts to the research case studies could benefit from this thesis' findings and recommendations. This suggests the need for similar studies to benefit medical schools with different contexts.

Another methodological limitation is the absence of perspectives from other stakeholders, such as students and regulatory bodies' leaders. Students' voices are not included since this study focuses on implementation rather than perception, and students had minimal input towards the OSCE's implementation. Although this keeps thesis research focused on addressing the research question, I recognised that some student input could be more valuable than I had initially anticipated. They would contribute to the study by describing their experiences with the OSCE, how they would evaluate the current OSCE, and what they expect from an effective and fair OSCE. Furthermore, gathering data from decision-makers in regulatory bodies, such as NCAAA leaders, would offer valuable additional insight. Their participation would clarify how these bodies assure the quality of medical school assessment. Expanding the range of stakeholders with whom this issue is discussed would add additional value to this project, so this can be the focus of future research.

8.5.2.2 Communication issues

In qualitative research, the accuracy of participants' spoken language during interviews and FGs is of utmost importance so that they can articulate themselves precisely. Despite the fact that they use English at work, the participants only speak it as a second language. I noticed some participants had difficulty articulating their views, which posed a challenge for this study. However, I was aware of this issue beforehand, so I was very cautious when interpreting their sentences. For example, participants may use phrases with nuanced meanings, so I rephrased their responses to confirm their accuracy. In addition, I was taking notes to provide commentary and capture the context in order to inform the analysis later. Another approach involved sharing portions of a transcribed and coded interview with my supervisors so they could provide feedback and review the meaning and accuracy of the codes generated.

Social desirability bias is another communication issue that may influence the findings of this study, as participants may alter their responses to please the researcher (Grimm, 2010). For example, some participants may focus on the positive aspects of their practice and avoid negative aspects. Consequently, I analysed the data with caution, as some responses may be imprecise (e.g., exaggerated), and some information may simply be concealed. This issue may add complexity to the interpretation of the findings; however, I attempted to mitigate this effect by triangulating the findings from multiple sources and participants. To further mitigate this effect, prior to the interviews, I endeavoured to establish a good rapport with the participant and notify them that there are no right or

wrong answers and that I am only interested in hearing about their current practice (Bergen and Labonté, 2020).

8.6 Reflexivity

From a constructivist philosophical stance, it is impossible to separate the research from the researcher who designed the methodology, decided on the methods, collected the data, analysed the data, and wrote the research (Bulpitt and Martin, 2010; Varpio et al., 2021). Therefore, the researcher's reflexivity is deemed a quality measure that helps readers better understand the research process (Dodgson, 2019). According to Walsh (2003) and Olmos-Vega et al. (2022), reflexivity is a multifaceted practice in which various aspects must be considered to provide a comprehensive overview of the researcher's impact on the research. They recommend discussing four areas to carry out rigorous qualitative research: personal, interpersonal, methodological, and contextual reflexivity.

8.6.1 Personal reflexivity

As a qualified medical doctor working in a Saudi medical school, I had previously only encountered positivist viewpoints and quantitative research methods. However, being a lecturer in a Medical Education Unit and holding a master's degree in medical education developed my perception of knowledge. My PhD journey shifted my ontological and epistemological stance towards a more constructivist paradigm. Throughout this journey, my thinking evolved, and I found it important to maintain a reflective journal in which I tracked my thoughts, experiences, assumptions, and ideas (Annink, 2017). I found this method helpful in making informed decisions and upholding transparency in my writing. This practice helped me to better understand my position as a researcher and made me realise that my interest in the OSCE long predates my PhD studies. My experience in the workplace and with other medical schools in Saudi Arabia enables me to confront the challenges of OSCE implementation. Therefore, I wrote a PhD proposal on this topic, which my supervisors deemed worthy of in-depth investigation.

At the outset of this investigation, I envisioned a more linear model in which I would use an operational guide to solve the problem. However, when I immersed myself in three data sources from two case studies, I discovered that the issue is more complex than I initially anticipated due to its educational, financial, and sociocultural facets. In fact, this further justifies the use of an in-depth educational case study design. The first round of data analysis and theme generation reflected my initial thoughts, but after several rounds of extensive data engagement, I developed the current themes that reflected my new insight. Although developing an OSCE operational guide is important, it is insufficient to address the other issues discussed in this thesis.

This is the most substantial piece of work I have ever completed, requiring a significant personal commitment and investment. However, my PhD journey has been challenging and illuminating; I had numerous “aha!” moments that altered my thinking, reshaping me into a different but better version of myself. Another unique experience for me has been writing subnational research in a second language. Arabic is my first language, and it is very different from English. They differ in many ways, but perhaps the most significant is their approach to logic and sentence structure. Therefore, I made every effort to improve my academic English by reading and writing extensively and by attending academic writing workshops. Although this learning journey will never end, I believe incorporating my supervisors’ feedback and suggestions have greatly improved my writing style. I learned that good research is an iterative process of re-writing and re-editing. So, I did this throughout the thesis, for example, deciding which ideas and quotes should be included and which are irrelevant to my work. This is a time-consuming and intellectually demanding process. However, I believe this is the best way to write a thesis.

I had three opportunities to present this research at international conferences, which advanced my knowledge and allowed me to network with the medical education community. While I benefited from audience questions and feedback, I observed that how important this topic is to many medical schools worldwide. My future goals include research and publication in medical education. I intend to create a network of young Saudi medical educators and schedule regular meetings at which we can discuss local issues and field advancements. This group can collaborate to publish research and form an advisory group for medical schools akin to AMEE and other organisations' early career researchers’ support groups.

8.6.2 Interpersonal reflexivity

I am an ‘insider researcher’ since I share the same culture as the participants and am familiar with the study’s community. This could influence the research process and findings due to my familiarity with the context, which might limit my perspective and cause me to overlook alternative viewpoints (Walsh, 2003; Barrett et al., 2020). However, I acknowledged my status as a quasi-ethnographer and subject-matter specialist. Consequently, I managed my tension while collecting and analysing the data. I kept a clear distance from what they told me, what I knew about them, and what I heard elsewhere (Olmos-Vega et al., 2022). I was self-reflective by maintaining a self-reflective journal throughout the analysis and being transparent and forthcoming throughout the writing (Berger, 2015; Sudirman et al., 2021). For example, I identified all that is important by looking at the codes, themes, and interpretations and discussing them with the supervisors. This, along with the reflective journal, has afforded me the opportunity to allow them to try and help me maintain that distance.

Another issue is that some participants know me, so this relationship might influence my interaction with them (Cohen et al., 2018). However, I was aware of my situation; thus, I was reflexive to myself and careful that it did not affect the data. For example, I constantly asked myself whether this dynamic relationship influences participants' responses to highlight more negative or positive responses (Bulpitt and Martin, 2010). I was aware of this type of social desirability bias, for example, when participants attempted to please me as a researcher (Grimm, 2010; Varpio et al., 2021). However, our previous relationship had no power or authority over one another. Therefore, I felt they were honest, cooperative, and willing to share their real experiences.

Being familiar to some participants made recruitment relatively easier and facilitated a cordial, productive discussion. This paved the way for easy communication with them and boosted my confidence during the discussion. Some of them, however, misunderstood me because they knew my background and speciality, and they wondered why I would ask questions when the answers were obvious to both them and me. For instance, in some FGs, when I asked if there were discrepancies between examiners in the parallel OSCEs, some participants responded, "Do you not recognise that?". Others attempted to investigate why I did not recognise such a fact. I was surprised by their response, so I took a step back and explained that I ask questions regardless of my position; what I really require are your thoughts and how you deal with things (Cohen et al., 2018). This was most likely due to their unfamiliarity with the nature of the qualitative approach in general.

Informal conversations before and after the recorded meetings contribute significantly to my understanding of these medical schools (Annink, 2017). During those off-the-record discussions, the faculty discussed things I had not expected them to discuss, yet it reflected the dilemmas in their OSCE implementation. They almost wanted that meeting to be a consultation session, implying their need for assessment specialists' advice in their practice.

8.6.3 Methodological reflexivity

As a qualitative researcher, I embrace subjectivity as part of my constructivist philosophical stance. Consequently, I have given careful consideration to this paradigmatic stance throughout the thesis. Even though I am the sole researcher for this project, discussing the methodology with supervisors, transfer examiners, and colleagues was beneficial. For example, I had planned to conduct a mixed-method study incorporating quantitative data by conducting a psychometric study analysing students' OSCE results. However, the transfer panel recommended that this step be kept as a follow-up post-doctoral study. They argue that the proposed qualitative part will suffice for the purposes of a PhD project and that any additional work will render the thesis unfeasible, given that I am a single researcher with a limited timeframe. At the end of

this project, I found that the advice was useful; qualitative research is difficult and time-consuming, but it yields rich data and an in-depth understanding of the research questions.

The process of collecting the documents and conducting the interviews and FGs added a wealth of understanding about the institutional culture of participating medical schools, how to approach academics and communicate effectively. For example, I learned that approaching the gatekeepers with appreciation and perseverance is essential to this process. The data collection process is an important aspect of understanding the context that would influence the data analysis and findings. Using the Codebook and RTA approaches to analyse the data allowed me to immerse myself in the data during the transcription, familiarisation, coding, theme generation, theme review, and writing-up stages. At each stage, I gained a deeper level of immersion into the data, to the point where I noticed my perception of the data had markedly developed.

Reflecting on my methodological approach strengthened my new belief that qualitative methods are essential for assessment research, which has traditionally been dominated by quantitative research (Hodges, 2003a). I would therefore recommend qualitative methods to other researchers. I find that the combination of analysing assessment documents and conducting interviews and FGs is a useful methodology for other researchers in the field. However, if I were to repeat this study, I would collect data using additional qualitative methods and include more stakeholders. For example, the ethnographic approach would enable me to observe the actual OSCE implementation as it takes place on the ground, which will provide valuable input to the research. Moreover, considering other stakeholders' perceptions, such as students and policymakers in accreditation agencies, would further develop the findings and be a useful post-doctoral focus. There are always opportunities to conduct follow-up research to further develop the work, for example, by incorporating quantitative approaches such as psychometric studies, as I intend to do soon after completing this thesis (8.7).

8.6.4 Contextual reflexivity

As an insider researcher, I am able to comprehend the culture and environment of these medical schools. In a similar context, I have been examined by the OSCE as a student, and I have also served as an OSCE designer and examiner for multiple OSCEs. I have worked with these medical schools in many workshops and local conferences and heard stories about their practices, all of which contribute to a rich conceptualisation of their OSCE implementation process. While this allows me to gain a deeper understanding of the issues under investigation, it may cause me to overlook things I have taken for granted due to my familiarity with the context. However, I took field notes to document many of these issues to make myself conscious of them. For example, I took notes on the funding source and how it affects decision-making. Furthermore, I have been

transparent about the context of each medical school, enabling the reader to understand and evaluate the findings. Therefore, my supervisors served as a safety net for any unconscious confirmation bias I may have had.

While gathering the data, I occasionally heard about suboptimal OSCE implementation practices that could negatively impact students and other stakeholders. These practices would undermine the credibility of their OSCE. As a researcher, I was not in a position to advise them of these practices during the data collection period (Othman and Hamid, 2018). I found this a difficult dilemma, and after reflection and discussion with my supervisors, I have decided to offer a supplementary report based on this thesis to inform participants about OSCE implementation best practises. This would partially reward the participants for their time participating in this study and help them improve their OSCE implementation.

As discussed in the Methodology Chapter, I purposefully selected medical schools for my case studies. Thus, I was aware of the contextual distinctions between the two. Although they share many similarities, they also have some differences. Working with newly established medical schools with different business models led me to discover intriguing findings. The process of conducting this PhD in these specific contexts took me beyond my initial expectations, revealing things to me throughout the course of this research. For example, I started to see different challenges and opportunities in each medical school and look for change theories, which are things that I did not start with but evolved during my PhD journey. Although those issues were not explicit in my research questions, I cannot ignore them. This research increases my awareness of context-specific nuances and their influence on practice. As a result, I intend to expand on this work by looking for ways to bring change to medical schools during the government's university privatisation movement.

8.7 Next steps

Given my current understanding of the assessment literature and current OSCE practice in Saudi, I consider myself a change agent (Bartunek, 2014). I will employ the skills I gained during my doctoral studies to improve medical education in my context. By being so immersed in this work for so long, I can now support other medical schools' assessments, support faculty development, encourage collaborative networking, and bring together the regulators to help develop a change roadmap collaboratively.

I also consider this exploratory study a starting point for further assessment and OSCE implementation research. For example, I plan to publish parts of this research to inform medical schools of similar contexts, as I believe it would greatly benefit their practice. Additionally, I am motivated to conduct a follow-up study to this thesis to examine the findings of this research by conducting an empirical quantitative study that analyses OSCE data psychometrically. I would also conduct additional research to examine the

perspectives of other stockholders, such as students and regulatory bodies' decision-makers. Their perspective would contribute to a better understanding of the OSCE's implementation. I am keen to extend this methodology further to take an additional ethnographic focus by observing how different things play out in a real OSCE (Kearney et al., 2018). An ethnographic observation can collect data from on-site observation for pre-OSCE preparation, attending live OSCE stations, and participating in post-OSCE discussions to better understand its implementation (Cohen et al., 2018). The findings of such studies will be valuable because they can either support, refute, or add a new perspective to the current research.

References

Abdelaziz, A., Hany, M., Atwa, H., Talaat, W. and Hosny, S. 2016. Development, implementation, and evaluation of an integrated multidisciplinary Objective Structured Clinical Examination (OSCE) in primary health care settings within limited resources. *Med Teach.* **38**(3), pp.272-279.

Abdulghani, H.M., Ponnamperuma, G. and Amin, Z. 2015. *An Essential Guide To Developing, Implementing, And Evaluating Objective Structured Clinical Examination (OSCE).* 1st ed. Singapore: World Scientific.

Abdullatif, A. 1992. An examination of the examinations: the reliability of the objective structured clinical examination and clinical examination. *Med Teach.* **14**(2-3), pp.179-183.

Abdulrahman, K.A.B. 2012. *The value of medical education research in Saudi Arabia*. Taylor & Francis. 34. pp.S1-S3.

Abrol, M. 2016. Emerging trends of privatization of education in India. *International Journal of Educational Administration.* **8**(1), pp.1-6.

Abu-Zaid, A., Salem, H. and Alkattan, K. 2020. The Saudi medical licensure examination-clinical skills (SMLE-CS): a call for Implementation. *Journal of Family Medicine and Primary Care.* **9**(1), pp.12-15.

ACCLAiM. 2023. *Australasian Collaboration for Clinical Assessment in Medicine — ACCLAiM.* [Online]. [Accessed 31 July]. Available from: https://acclaim.edu.au/

Al-Eidan, F., Baig, L.A., Magzoub, M.E. and Omair, A. 2016. Reliability and validity of the faculty evaluation instrument used at king saud bin abdulaziz university for health sciences: Results from the haematology course. *Journal of the Pakistan Medical Association.* **66**(4), pp.453-457.

Al-Haddad, M.K. 2009. Use of OSCE in undergraduate assessment of Psychiatry at the College of Medicine and Medical Sciences - Arabian Gulf University, Bahrain. *Journal of the Bahrain Medical Society.* **21**(1), pp.221-222.

Al-Hashimi, K., Said, U.N. and Khan, T.N. 2023. Formative Objective Structured Clinical Examinations (OSCEs) as an Assessment Tool in UK Undergraduate Medical Education: A Review of Its Utility. *Cureus.* **15**(5), pp.1-7.

Al-Mously, N., Salem, R. and Al-Hamdan, N. 2013. The impact of gender and English language on the academic performance of students: An experience from new Saudi medical school. *Journal of Contemporary Medical Education.* **1**(3), pp.170-176.

Al-Naami. 2008. Reliability, validity, and feasibility of the Objective Structured Clinical Examination in assessing clinical skills of final year surgical clerkship. *Saudi Med J.* **29**(12), pp.1802-1807.

Al-Naami, El-Tinay, O.F., Khairy, G.A., Mofti, S.S. and Anjum, M.N. 2011. Improvement of psychometric properties of the objective structured clinical examination when assessing problem solving skills of surgical clerkship. *Saudi Med J.* **32**(3), pp.300-304.

Al-Rabeeah, A. 2003. The history of health care in the Kingdom of Saudi Arabia with emphasis on pediatric surgery. *Saudi medical journal.* **24**, pp.S9-10.

Al-Rubaish, A.M., Al-Umran, K.U. and Wosornu, L. 2005. An audit of assessment tools in a medical school in eastern saudi arabia. *J Family Community Med.* **12**(2), pp.101-105.

Al-Wassia, R., Hamed, O., Al-Wassia, H., Alafari, R. and Jamjoom, R. 2015. Cultural challenges to implementation of formative assessment in Saudi Arabia: An exploratory study. *Medical Teacher.* **37**(1), pp.S9-S19.

Al Shawwa, L.A. 2012. The establishment and roles of the Medical Education Department in the faculty of Medicine, King Abdul Aziz University, Jeddah Saudi Arabia. *Oman Medical Journal.* **27**(1), pp.4-9.

Alaidarous, S., Mohamed, T.A., Masuadi, E., Wali, S. and AlMalki, A. 2016. Saudi Internal Medicine Residents' Perceptions of the Objective Structured Clinical Examination as a Formative Assessment Tool. *Health Professions Education.* **2**(2), pp.121-129.

Alamro, A.S., Alghasham, A.A., Al-Shobaili, H.A., Alhomaidan, H.T., Salem, T.A., Wadi, M.M. and Saleh, M.N. 2023. 10 years of experience in adopting, implementing and evaluating progress testing for Saudi medical students. *Journal of Taibah University Medical Sciences.* **18**(1), p175.

Alghamdi, H.H. 2019. Improving the competitiveness of emerging universities according to the requirements of privatizing higher education in the Kingdom of Saudi Arabia: a proposed scenario. *The specialized international educational journal.* **8**(9), pp.84-97.

Alghamdi, K.S., Katib, B.H., Alhoqail, A.A. and Al-khatib, T.A. 2016. Perception and acceptance of senior medical students at King Abdulaziz University of the use of Objective Structured Clinical Examination as a tool of assessment. *MedEdPublish.* **5**(2), pp.1-9.

Alghasham, A.A. 2012. Effect of students' learning styles on classroom performance in problem-based learning. *Medical Teacher.* **34**(1), pp.S14-S19.

Aljohani, K.A.S. 2020. Nursing education in Saudi Arabia: History and development. *Cureus.* **12**(4), pp.1-6.

Alkhateeb, N., Salih, A.M., Shabila, N. and Al-Dabbagh, A. 2022. Objective structured clinical examination: Challenges and opportunities from students' perspective. *Plos one.* **17**(9), pe0274055.

Allan, D.D.A., Merrill, E., Allen, P., Armstrong, M.L. and Merritt, J. 2023. Recommendations After an Objective Structured Clinical Examination Program Evaluation. *The Journal of Continuing Education in Nursing.* **54**(4), pp.169-175.

Allan, G.A. and Skinner, C.J. 1991. *Handbook for Research Students in the Social Sciences.* London: The falmer Press.

Almasabi, M. 2013. An overview of health system in Saudi Arabia. *Res J Med Sci.* **7**(3), pp.70-74.

Almisnid, K.M. 2016. *Differences in learning strategies and study skills between high and low achievers among undergraduate pre-clinical medical students at a medical college: A qualitative research*. Master thesis, Cardiff University.

Alnahdi, M.A., Alhaider, A., Bahanan, F., Aldubaikhi, A., Aljehani, A., Omair, A. and Alaqeel, M. 2021. The impact of the English medical curriculum on medical history taking from Arabic speaking patients by medical students. *Journal of Family Medicine and Primary Care.* **10**(3), pp.1425-1430.

Alnasser, S.S.A. 2016. *It does not need to be voiced to be counted. Non-verbal behaviour influences assessors' global marking when examining medical students using objective structured clinical examinations*. thesis, University of Leeds.

Alnufaiee, S.J.-A. 2019. The privatization of Saudi universities in the light of indicators of knowledge economy and national transformation. *Journal of Scientific Research in Education.* **20**(Section 10), pp.251-306.

Alomair, A. 2021. *The Feedback Experience of Saudi Arabian Medical Students and Tutors in PBL Medical Curricula: A Mixed Method Study*. thesis, University of Sheffield.

Alrebish, S.A. 2014. *The Impact of an Accreditation System on the Quality of Undergraduate Medical Education in Saudi Arabia*. thesis, Monash University.

Alrehaily, A., Alharbi, N., Zaini, R. and AlRumayyan, A. 2022. Perspectives of the key stakeholders of the alignment and integration of the SaudiMEDS framework into the Saudi medical licensure examination: a qualitative study. *Advances in Medical Education and Practice.* **13**, pp.59-69.

Alrubai, Z.A. 2011. Privatization of university education in Jordan: a field study. *Journal of the College of Education in Ismailia.* **21**, pp.91-134.

Alsaid, A.H. and Al-Sheikh, M. 2017. Student and faculty perception of objective structured clinical examination: A teaching hospital experience. *Saudi journal of medicine & medical sciences.* **5**(1), pp.49-55.

Alsharif, N.Z. 2011. *Faculty members' enthusiasm: a blessing or a curse?* : American Journal of Pharmaceutical Education. 75. pp.23-24.

Alsulimani, L.K., Al-Otaiby, F.M., Alnofaiey, Y.H., Binobaid, F.A., Jafarah, L.M. and Khalil, D.A. 2020. Attitudes Towards Introduction of Multiple Modalities of Simulation in Objective Structured Clinical Examination (OSCE) of Emergency Medicine (EM) Final Board Examination: A Cross-Sectional Study. *Open Access Emergency Medicine.* **12**, pp.441-449.

Altbach, P.G. and Levy, D.C. 2005. *Private Higher Education: A Global Revolution.* Rotterdam: Sense Publishers.

Althubaiti, M.b.O. 2020. Privatization of university education in the light of the Kingdom of Saudi Arabia's vision 2030: a proposed vision. *Journal of Educational and Psychological Sciences.* **14**(2), pp.860-828.

Altshuler, L. and Kachur, E. 2001. A culture OSCE: teaching residents to bridge different worlds. *Academic Medicine.* **76**(5), p514.

Alzahrani, A.b.A.Q.A. 2019. Obstacles to the privatization of Saudi universities in the light of Vision 2030 from the point of view of academic leaders. *The specialized international educational journal.* **8**(11), pp.158-136.

AMEE. 2023. *Association For Medical Education.* [Online]. [Accessed 29 July]. Available from: https://amee.org/awards-prizes/aspire-award

Amin, Kaliyadan, F. and Al-Muhaidib, N.S. 2011. Medical students' assessment preferences at King Faisal University, Saudi Arabia. *Adv Med Educ Pract.* **2**, pp.95-103.

Amin, Z., Chong, Y.S. and Khoo, H.E. 2006. *Practical guide to medical student assessment.* Singapore: World Scientific.

Amr, M. and Amin, T. 2012. Assessment methods of an undergraduate psychiatry course at a saudi university. *Sultan Qaboos Univ Med J.* **12**(2), pp.214-220.

Anderson, S., Allen, P., Peckham, S. and Goodwin, N. 2008. Asking the right questions: scoping studies in the commissioning of research on the organisation and delivery of health services. *Health research policy and systems.* **6**, pp.7-7.

Annink, A. 2017. Using the research journal during qualitative data collection in a cross-cultural context. *Entrepreneurship Research Journal.* **7**(1), pp.223-240.

Ansari, T., Al Mansour, M. and Tanzil, S. 2021. Acceptability of objective structured clinical examination (Osce) as an effective assessment tool among

undergraduate medical students of majmaah university, ksa. *Rawal Medical Journal.* **46**(3), pp.685-689.

Arain, F.R. 2021. Does test anxiety and gender have an impact on OSCE Performance among medical students? *World Family Medicine.* **19**(11), pp.101-107.

Arksey, H. and O'Malley, L. 2005. Scoping studies: Towards a Methodological Framework. *International Journal Social Research Methodology.* **8**(1), pp.19-32.

Armstrong, R., Hall, B.J., Doyle, J. and Waters, E. 2011. 'Scoping the scope' of a cochrane review. *Journal of Public Health.* **33**(1), pp.147-150.

Arthur, J., Waring, M., Coe, R. and Hedges, L. 2012. *Research Methods and Methodologies in Education.* London, UK: SAGE Publications.

ASME. 2023. *Psychometric, Measurement, and Quality in Assessment Special Interest Group of ASME.* [Online]. [Accessed 27 July]. Available from: https://www.asme.org.uk/psychometrics/

Ataro, G., Worku, S. and Asaminew, T. 2020. Experience and challenges of objective structured clinical examination (OSCE): perspective of students and examiners in a clinical department of Ethiopian University. *Ethiopian Journal of Health Sciences.* **30**(3), pp.417-426.

Attia, J. 2022. *Signs of renaissance, the development of the Saudi health system..the roles of pioneers.* [Online]. [Accessed 31 July 2023]. Available from: https://wtn.sa/a/1107996

Attride-Stirling, J. 2001. Thematic networks: an analytic tool for qualitative research. *Qualitative research.* **1**(3), pp.385-405.

Ayuob, N.N., Bahumdain, S.Z., AL-Najei, Z.A., Khobrani, F.M. and El Deek, B.S. 2015. Becoming a doctor, what is behind this decision? A comparative study between governmental and private medical school in Saudi Arabia. *J Contemp Med Edu.* **3**(4), p155.

Badawi, M.A., Hussain, H.O., Badawi, M.A., Al Shawwa, L.A., Jamjoom, R.A., Park, Y.S. and Tekian, A. 2023. Learning transfusion medicine through scoring objective structured clinical examination. *Transfusion.* **63**(S1), pp.S20-S27.

Badyal, D.K. and Sharma, M. 2020. Internal assessment in new MBBS curriculum: Methods and logistics. *International Journal of Applied and Basic Medical Research.* **10**(2), pp.68-75.

Bajammal, S., Zaini, R., Abuznadah, W., Al-Rukban, M., Aly, S.M., Boker, A., Al-Zalabani, A., Al-Omran, M., Al-Habib, A., Al-Sheikh, M., Al-Sultan, M., Fida, N., Alzahrani, K., Hamad, B., Al Shehri, M., Abdulrahman, K.B., Al-Damegh, S., Al-Nozha, M.M. and Donnon, T. 2008. The need for national medical licensing examination in Saudi Arabia. *BMC Medical Education.* **8**(53), pp.1-15.

Bakhsh, T., Al Kashgari, R.H., Sibiany, A.M., Al Mashat, F.M., Meccawy, A.A., Al Thubaity, F.K. and Johary, A.A. 2005. Running OSCE for a Large Group of Students: Experience of the Surgical Department King Abdulaziz University, Faculty of Medicine. *Journal of King Abdulaziz University-Medical Sciences.* **12**, pp.43-52.

Bakhsh, T., Sibiany, A., Al-Mashat, F., Meccawy, A. and Al-Thubaity, F. 2009. Comparison of students' performance in the traditional oral clinical examination and the objective structured clinical examination. *The Saudi Medical Journal.* **30**(4), pp.555-557.

Baloh, J., Zhu, X. and Ward, M.M. 2018. Implementing team huddles in small rural hospitals: How does the Kotter model of change apply? *Journal of nursing management.* **26**(5), pp.571-578.

Banihashemi, S. 2011. The role of communication to improve organizational process. *European Journal of Humanities and Social Sciences.* **1**(1), pp.13-24.

Bank, L., Jippes, M., Leppink, J., Scherpbier, A.J., den Rooyen, C., van Luijk, S.J. and Scheele, F. 2017. Are they ready? Organizational readiness for change among clinical teaching teams. *Advances in medical education and practice.* pp.807-815.

Barman, A. 2005. Critiques on the objective structured clinical examination. *Annals-Academy of Medicine Singapore.* **34**(8), p478.

Barrett, A., Kajamaa, A. and Johnston, J. 2020. How to… be reflexive when conducting qualitative research. *The clinical teacher.* **17**(1), pp.9-12.

Bartman, I., Smee, S. and Roy, M. 2013. A method for identifying extreme OSCE examiners. *The clinical teacher.* **10**(1), pp.27-31.

Bartunek, J.M. 2014. *Organizational and Educational Change: The Life and Role of A Change Agent Group.* Nwe York: Taylor & Francis.

Beach, Y. 2016. The Socio-political dynamics and complexity of organisational change projects: A research agenda. In: *Project Managemement Development-Practice and Perspectives*: Sheffield, pp.20-30.

Bearman, M., Ajjawi, R., Bennett, S. and Boud, D. 2020. The hidden labours of designing the Objective Structured Clinical Examination: a Practice Theory study. *Advances in Health Sciences Education.* **26**, pp.637-651.

Bearman, M., Dawson, P., Bennett, S., Hall, M., Molloy, E., Boud, D. and Joughin, G. 2017. How university teachers design assessments: a cross-disciplinary study. *Higher Education.* **74**, pp.49-64.

Bell, J. and Waters, S. 2014. *Doing Your Research Project: A Guide For First-Time Researchers.* 6th ed. Berkshire: McGraw-Hill Education.

Ben-David, M.F. 2000. AMEE Guide No. 18: Standard setting in student assessment. *Medical Teacher.* **22**(2), pp.120-130.

Ben-David, M.F., Davis, M.H., Harden, R.M., Howie, P.W., Ker, J. and Pippard, M.J. 2001. AMEE Medical Education Guide No. 24: Portfolios as a method of student assessment. *Medical Teacher.* **23**(6), pp.535-551.

Bennett, D., Barrett, A. and Helmich, E. 2019. How to… analyse qualitative data in different ways. *The clinical teacher.* **16**(1), pp.7-12.

Berendonk, C., Stalmeijer, R.E. and Schuwirth, L.W. 2013. Expertise in performance assessment: assessors' perspectives. *Advances in Health Sciences Education.* **18**, pp.559-571.

Bergen, N. and Labonté, R. 2020. "Everything is perfect, and we have no problems": detecting and limiting social desirability bias in qualitative research. *Qualitative health research.* **30**(5), pp.783-792.

Berger, R. 2015. Now I see it, now I don't: Researcher's position and reflexivity in qualitative research. *Qualitative research.* **15**(2), pp.219-234.

Bergman, E., de Feijter, J., Frambach, J., Godefrooij, M., Slootweg, I., Stalmeijer, R. and van der Zwet, J. 2012. AM Last Page: A Guide to Research Paradigms Relevant to Medical Education. *Academic Medicine.* **87**(4), p545.

Besar, M.N.A., Siraj, H.H., Abd Manap, R., Mahdy, Z.A., Yaman, M.N., Kamarudin, M.A. and Mohamad, N. 2012. Should a single clinician examiner be used in objective structure clinical examination? *Procedia-Social and Behavioral Sciences.* **60**, pp.443-449.

Bhatnagar, K., Saoji, V. and Banerjee, A. 2011. Objective structured clinical examination for undergraduates: Is it a feasible approach to standardized assessment in India? *Indian Journal of Ophthalmology.* **59**(3), pp.211-214.

Bin-Abdulrahman, K. 2011. Saudi Arabia does not need an Abraham Flexner. *Medical Teacher.* **33**(1), pp.74-75.

Bin-Abdulrahman, K., Harden, R. and Patricio, M. 2012. Medical education in Saudi Arabia: An exciting journey. *Medical Teacher.* **34**, pp.S4-S5.

Blaikie, N. and Priest, J. 2019. *Designing Social Research: The Logic of Anticipation.* 3ed ed. Cambridge, UK: Polit Press.

Boddy, C.R. 2016. Sample size for qualitative research. *Qualitative Market Research: An International Journal.* **19**(4), pp.426-432.

Bonvin, R., Nendaz, M., Frey, P., Schnabel, K.P., Huwendiek, S. and Schirlo, C. 2019. Looking back: twenty years of reforming undergraduate medical training and curriculum frameworks in Switzerland. *Gms Journal for Medical Education.* **36**(5), pp.1-17.

Boud, D. 2000. Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education.* **22**(2), pp.151-167.

Boud, D. 2015. Feedback: ensuring that it leads to enhanced learning. *Clinical Teacher.* **12**(1), pp.3-7.

Boud, D. and Molloy, E. 2013. *Feedback in Higher and Professional Education: Understanding it and Doing it Well.* Abingdon: Routledge.

Boud, D. and Soler, R. 2016. Sustainable assessment revisited. *Assessment & Evaluation in Higher Education.* **41**(3), pp.400-413.

Boulet, J.R. and McKinley, D.W. 2022. Criteria for a Good Assessment. In: Mcgaghie, W. ed. *International Best Practices for Evaluation in the Health Professions.* Florida: CRC Press.

Boursicot, K. 2010. Structured assessments of clinical competence. *British Journal of Hospital Medicine.* **71**(6), pp.342-344.

Boursicot, K., Etheridge, L., Setna, Z., Sturrock, A., Ker, J., Smee, S. and Sambandam, E. 2011. Performance in assessment: Consensus statement and recommendations from the Ottawa conference. *Medical Teacher.* **33**(5), pp.370-383.

Boursicot, K., Kemp, S. and Fuller, R. 2022. Quality Assurance of Objective Structured Clinical Examination. In: Malau-Aduli, B.S., et al. eds. *Understanding Assessment in Medical Education through Quality Assurance.* London: McGraw Hill, pp.68-94.

Boursicot, K., Kemp, S., Wilkinson, T., Findyartini, A., Canning, C., Cilliers, F. and Fuller, R. 2020. Performance assessment: Consensus statement and recommendations from the 2020 Ottawa Conference. *Medical Teacher.* **43**(1), pp.58-67.

Boursicot, K. and Roberts, T. 2005. How to set up an OSCE. *The clinical teacher.* **2**(1), pp.16-20.

Bowen, G.A. 2009. Document analysis as a qualitative research method. *Qualitative research journal.* **9**(2), pp.27-40.

Boyatzis, R.E. 1998. *Transforming Qualitative Information: Thematic Analysis and Code Development.* London: SAGE Publications.

Braier-Lorimer, D.A. and Warren-Miell, H. 2022. A peer-led mock OSCE improves student confidence for summative OSCE assessments in a traditional medical course. *Medical Teacher.* **44**(5), pp.535-540.

Brannick, M.T., Erol-Korkmaz, H.T. and Prewett, M. 2011. A systematic review of the reliability of objective structured clinical examination scores. *Medical Education.* **45**(12), pp.1181-1189.

Braun, V. and Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology.* **3**(2), pp.77-101.

Braun, V. and Clarke, V. 2013. *Successful Qualitative Research.* London: SAGE.

Braun, V. and Clarke, V. 2020. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative research in psychology.* **18**(3), pp.1-25.

Braun, V. and Clarke, V. 2021. Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and Psychotherapy Research.* **21**(1), pp.37-47.

Braun, V. and Clarke, V. 2022. *Thematic Analysis: A Practical Guide.* London: SAGE Publications.

Braun, V., Clarke, V., Hayfield, N. and Terry, G. 2019. Thematic Analysis. In: Liamputtong, P. ed. *Handbook of Research Methods in Health Social Sciences.* Singapore: Springer Singapore, pp.843-860.

Braun, V., Clarke, V. and Rance, N. 2014. How to Use Thematic Analysis with Interview Data. In: Vossler, A. and Moller, N. eds. *The Counselling & Psychotherapy Research Handbook.* London: Sage, pp.183–197.

Bridges, W. 1995. *Managing Transitions: Making The Most Of Change, 2nd Edition.* UK: Hachette Books.

Brits, H., Bezuidenhout, J., van der Merwe, L.J. and Joubert, G. 2020. Assessment practices in undergraduate clinics medicine raining: What do we do and how can we improve? *African Journal of Primary Health Care & Family Medicine.* **12**(1), pp.1-7.

Brits, H., Joubert, G., Bezuidenhout, J. and van der Merwe, L. 2021. Evaluation of assessment marks in the clinical years of an undergraduate medical training programme: Where are we and how can we improve? *African Journal of Health Professions Education.* **13**(4), pp.222-228.

Bryman, A. 2007. Effective leadership in higher education: A literature review. *Studies in higher education.* **32**(6), pp.693-710.

Bryman, A. 2016. *Social Research Methods.* 5th ed. Oxford: Oxford University Press.

Buchanan, D., Fitzgerald, L., Ketley, D., Gollop, R., Jones, J.L., Lamont, S.S., Neath, A. and Whitby, E. 2005. No going back: A review of the literature on sustaining organizational change. *International Journal of Management Reviews.* **7**(3), pp.189-205.

Bulpitt, H. and Martin, P.J. 2010. Who am I and what am I doing? Becoming a qualitative research interviewer. *Nurse researcher.* **17**(3), pp.7-16.

Burke, W.W. 2002. *Organization Change: Theory and Practice.* SAGE Publications.

Byrne, D. 2022. A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & quantity.* **56**(3), pp.1391-1412.

Campbell, K.A., Orr, E., Durepos, P., Nguyen, L., Li, L., Whitmore, C., Gehrke, P., Graham, L. and Jack, S.M. 2021. Reflexive thematic analysis for applied qualitative health research. *The Qualitative Report.* **26**(6), pp.2011-2028.

Carraccio, C. and Englander, R. 2000. The Objective Structured Clinical Examination: A Step in the Direction of Competency-Based Evaluation. *Archives of Pediatrics & Adolescent Medicine.* **154**(7), pp.736-741.

Case, S.M. and Swanson, D.B. 2001. *Constructing Written Test Questions For the Basic and Clinical Sciences.* [Online]. 3ed ed. United States of America: National Board of Medical Examiner. [Accessed 7 April 2016]. Available from: https://shorturl.at/auKO0

Casey, P.M., Goepfert, A.R., Espey, E.L., Hammoud, M.M., Kaczmarczyk, J.M., Katz, N.T., Neutens, J.J., Nuthalapaty, F.S., Peskin, E., Assoc Prof, G. and Obstet, U. 2009. To the point: reviews in medical education-the Objective Structured Clinical Examination. *American Journal of Obstetrics and Gynecology.* **200**(1), pp.25-34.

Chan, S.C.C., Choa, G., Kelly, J., Maru, D. and Rashid, M.A. 2023. Implementation of virtual OSCE in health professions education: A systematic review. *Medical Education.* (Preprint), pp.1-11.

Cho, E., Han, Y.M., Kang, Y., Kim, J.H., Shin, M.S., Oh, M., Cho, N.G., Jung, H.J. and Leem, J. 2022. Implementation of Objective Structured Clinical Examination on Diagnostic Musculoskeletal Ultrasonography Training in Undergraduate Traditional Korean Medicine Education: An Action Research. *Diagnostics.* **12**(7), p1707.

Chong, L., Taylor, S., Haywood, M., Adelstein, B.-A. and Shulruf, B. 2017. The sights and insights of examiners in objective structured clinical examinations. *Journal of educational evaluation for health professions.* **14**(34), pp.1-8.

Cleland, J. 2022. Exploring, measuring or both: considering the differences between qualitative, quantitative and mixed methods research. In: Cleland, J. and Durning, S.J. eds. *Researching Medical Education.* 2nd ed. Chichester: Wiley.

Cleland, J. and Durning, S.J. 2022. *Researching Medical Education.* 2nd ed. Chichester: Wiley.

Cleland, J. and Roberts, T.E. 2021. Medical education research. In: Dent, J., et al. eds. *A Practical Guide for Medical Teachers, E-Book: A Practical Guide for Medical Teachers, E-Book.* 6th ed. London: Elsevier Health Sciences.

Cleland, J.A., Abe, K. and Rethans, J.-J. 2009. The use of simulated patients in medical education: AMEE Guide No 42. *Medical teacher.* **31**(6), pp.477-486.

Cohen, D.J. and Crabtree, B.F. 2008. Evaluative criteria for qualitative research in health care: controversies and recommendations. *The Annals of Family Medicine.* **6**(4), pp.331-339.

Cohen, L., Manion, L. and Morrison, K. 2018. *Research Methods in Education.* 8th ed. Abingdon: Routledge.

Colbert-Getz, J.M. and Shea, J.A. 2021. Three key issues for determining competence in a system of assessment. *Medical teacher.* **43**(7), pp.853-855.

Collins, P. and Harden, R. 1998. AMEE Medical Education Guide No. 13: real patients, simulated patients and simulators in clinical examinations. *Medical Teacher.* **20**(6), pp.508-521.

Colliver, J.A. 2002. Educational Theory and Medical Education Practice: A Cautionary Note for Medical School Faculty. *Academic Medicine.* **77**(12), pp.1217-1220.

Connell, R. 2014. Using southern theory: Decolonizing social thought in theory, research and application. *Planning Theory.* **13**(2), pp.210-223.

Cook, D.A. 2014. When I say… validity. *Medical Education.* **48**, pp.948-949.

Cook, D.A., Zendejas, B., Hamstra, S.J., Hatala, R. and Brydges, R. 2014. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Advances in Health Sciences Education.* **19**(2), pp.233-250.

Cottrell, S. 2006. A matter of explanation: assessment, scholarship of teaching and their disconnect with theoretical development. *Medical Teacher.* **28**(4), pp.305-308.

Crabtree, B.F. and Miller, W.L. 1999. *Doing Qualitative Research.* 2nd ed. London: SAGE Publications.

Creswell, J.W. 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* 4th ed. London: SAGE.

Cristancho, S., Field, E. and Lingard, L. 2019. What is the state of complexity science in medical education research? *Medical education.* **53**(1), pp.95-104.

Cristancho, S., LaDonna, K. and Field, E. 2022. Visual methods in health professions research: purpose, challenges and opportunities. In: Cleland, J. and Durning, S.J. eds. *Researching Medical Education.* 2nd ed. Chichester: Wiley.

Cusimano, M.D., Cohen, R., Tucker, W., Murnaghan, J., Kodama, R. and Reznick, R. 1994. A comparative analysis of the costs of administration of an OSCE (objective structured clinical examination). *Academic Medicine.* **69**(7), pp.571-576.

Dadgar, S.R., Saleh, A., Bahador, H. and Baradaran, H.R. 2008. OSCE as a tool for evaluation of practical semiology in comparison to MCQ & Oral examination. *Journal of the Pakistan Medical Association.* **58**(9), pp.506-507.

Daniels, V.J., Ortiz, S., Sandhu, G., Lai, H., Yoon, M.N., Bulut, O. and Hillier, T. 2021. Effect of Detailed OSCE Score Reporting on Learning and Anxiety in Medical School. *J Med Educ Curric Dev.* **8**, pp.1-8.

Daniels, V.J. and Pugh, D. 2018. Twelve tips for developing an OSCE that measures what you want. *Medical teacher.* **40**(12), pp.1208-1213.

Daudt, H.M.L., van Mossel, C. and Scott, S.J. 2013. Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. *BMC Medical Research Methodology.* **13**(1), p48.

Davies, J.A. 2023. In search of learning-focused feedback practices: a linguistic analysis of higher education feedback policy. *Assessment & Evaluation in Higher Education.* (ahead-of-print), pp.1-15.

Davis, K., Drey, N. and Gould, D.J.I.J.N.S. 2009. What are scoping studies? A review of the nursing literature. *International Journal of Nursing Studies.* **46**(10), pp.1386-1400.

Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D. and Molloy, E. 2019. What makes for effective feedback: Staff and student perspectives. *Assessment & Evaluation in Higher Education.* **44**(1), pp.25-36.

Dent, J. and Harden, R.M. 2013. *A Practical Guide for Medical Teachers E-Book.* 4th ed. London: Elsevier Health Sciences.

Dent, J., Harden, R.M. and Hunt, D. 2021. *A practical guide for medical teachers.* Elsevier Health Sciences.

Denzin, N.K. and Lincoln, Y.S. 2018. *The SAGE Handbook of Qualitative Research.* 5th ed. London: SAGE.

Diab, M.I., Nasr, Z.G., El-Hajj, M.S., Elewa, H., El-Geed, H.A. and Wilby, K.J. 2019. Exploring the Influence of Language on Assessment Given a Mismatch Between Language of Instruction and Language of Practice. *Simulation in Healthcare.* **14**(4), pp.271-275.

Dodgson, J.E. 2019. Reflexivity in qualitative research. *Journal of Human Lactation.* **35**(2), pp.220-222.

Donohoe, C.L., Reilly, F., Donnelly, S. and Cahill, R.A. 2020. Is There Variability in Scoring of Student Surgical OSCE Performance Based on Examiner Experience and Expertise? *Journal of Surgical Education.* **77**(5), pp.1202-1210.

Downing, S.M. 2003. Validity: on the meaningful interpretation of assessment data. *Medical Education* **37**, pp.830–837.

Downing, S.M. 2006. Twelve steps for effective test development. In: Downing, S. and Haladyna, T. eds. *Handbook of test development.* United States: Taylor & Francis, p.778.

Downing, S.M. and Haladyna, T.M. 2004. Validity threats: overcoming interference with proposed interpretations of assessment data. *Medical education.* **38**(3), pp.327-333.

Duerson, M.C., Romrell, L.J. and Stevens, C.B. 2000. Impacting faculty teaching and student performance: nine years' experience with the Objective Structured Clinical Examination. *Teaching and Learning in Medicine.* **12**(4), pp.176-182.

Dunford, R., Cuganesan, S., Grant, D., Palmer, I., Beaumont, R. and Steele, C. 2013. “Flexibility” as the rationale for organizational change: A discourse perspective. *Journal of Organizational Change Management.* **26**(1), pp.83-97.

Dwivedi, N.R., Vijayashankar, N.P., Hansda, M., Dubey, A.K., Nwachukwu, F., Curran, V. and Jillwin, J. 2020. Comparing standard setting methods for objective structured clinical examinations in a Caribbean medical school. *Journal of Medical Education and Curricular Development.* **7**, pp.1-10.

Dyrvig, A.K., Kidholm, K., Gerke, O. and Vondeling, H. 2014. Checklists for external validity: a systematic review. *Journal of evaluation in clinical practice.* **20**(6), pp.857-864.

Elfaki and Al-Humayed, S. 2016. Medical students' perception of OSCE at the Department of Internal Medicine, College of Medicine, King Khalid University, Abha, KSA. *Journal of the College of Physicians and Surgeons Pakistan.* **26**(2), pp.158-159.

Elfaki, Badrek, A.H. and Jastaniah, S.A. 2008. Objective Structured Clinical Examination (OSCE) Seven Years Experience From Department Of Surgery, Faculty Of Medicine, Umm-Alqura, Makkah, Saudi Arabia. *Sudan Journal of Medical Sciences.* **3**(3), pp.185-190.

Elliott, R., Fischer, C.T. and Rennie, D.L. 1999. Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British journal of clinical psychology.* **38**(3), pp.215-229.

Engelkemeyer, S.W. 1998. Applying benchmarking in higher education: A review of three case studies. *Quality Management Journal.* **5**(4), pp.23-31.

Entwistle, N.J. and Ramsden, P. 2015. *Understanding Student Learning.* London: Routledge Revivals.

Epstein, R.M. 2007. Assessment in Medical Education. *New England Journal of Medicine.* **356**(4), pp.387-396.

ETEC. 2022. *Accreditation Handbook.* [Online]. [Accessed 29 July]. Available from: http://dl.icdst.org/pdfs/files/06072e0030ba10087ceb6c5a711fd2a2.pdf

ETEC. 2023a. *Education and Training Evaluation Commission.* [Online]. [Accessed 24 June]. Available from: https://shorturl.at/lwxLO

ETEC. 2023b. *Education and Training Evaluation Commission: Accreditation Handbook.* [Online]. [Accessed 29 July]. Available from: https://etec.gov.sa/service/Program%20Accreditation

Etheridge, L. and Boursicot, K. 2021. Performance and workplace assessment. In: Dent, J., et al. eds. *A Practical Guide for Medical Teachers, E-Book: A Practical Guide for Medical Teachers, E-Book.* 6th ed. London: Elsevier Health Sciences.

Eva, K.W., Bordage, G., Campbell, C., Galbraith, R., Ginsburg, S., Holmboe, E. and Regehr, G. 2016. Towards a program of assessment for health professionals: from training into practice. *Advances in Health Sciences Education.* **21**, pp.897-913.

Faherty, A., Counihan, T., Kropmans, T. and Finn, Y. 2020. Inter-rater reliability in clinical assessments: do examiner pairings influence candidate ratings? *BMC Medical Education.* **20**, pp.1-9.

Farghali, A.A. 2013. *Privatization of Academic Programs in Higher Education Institutions: An Exploratory Study in the Kingdom of Saudi Arabia*. thesis, Taibah University.

Farghaly, A. 2018. Comparing and Contrasting Quantitative and Qualitative Research Approaches in Education: The Peculiar Situation of Medical Education. *Education in Medicine Journal.* **10**(1), pp.3-11.

Finlay, L. 2006. 'Rigour','ethical integrity'or 'artistry'? Reflexively reviewing criteria for evaluating qualitative research. *British Journal of Occupational Therapy.* **69**(7), pp.319-326.

Flick, U. 2014. *The SAGE Handbook of Qualitative Data Analysis.* London: SAGE.

Flick, U. 2019. *An Introduction to Qualitative Research.* 6th ed. London: SAGE.

Flood, M., Strawbridge, J., Sheachnasaigh, E.N., Ryan, T., Sahm, L.J., Fleming, A. and Barlow, J.W. 2022. Supporting pharmacy students' preparation for an entry-to-practice OSCE using video cases. *Currents in Pharmacy Teaching and Learning.* **14**(12), pp.1525-1534.

Freeman, R. and Maybin, J. 2011. Documents, practices and policy. *Evidence & Policy: A Journal of Research, Debate and Practice.* **7**(2), pp.155-170.

Fuller, M., Henderson, S. and Bustamante, R. 2015. Assessment leaders' perspectives of institutional cultures of assessment: A Delphi study. *Assessment & Evaluation in Higher Education.* **40**(3), pp.331-351.

Fuller, R., Homer, M. and Pell, G. 2013. Longitudinal interrelationships of OSCE station level analyses, quality improvement and overall reliability. *Medical Teacher.* **35**(6), pp.515-517.

Fuller, R., Homer, M., Pell, G. and Hallam, J. 2017. Managing extremes of assessor judgment within the OSCE. *Medical teacher.* **39**(1), pp.58-66.

Fuller, R., Pell, G., Homer, M. and Roberts, T. 2012. Comments on 'A systematic review of the reliability of objective structured clinical examination scores'. *Medical Education.* **46**(3), pp.337-337.

Furman, G.E. 2008. The Role of Standardized Patient and Trainer Training in Quality Assurance for a High-Stakes Clinical Skills Examination. *The Kaohsiung Journal of Medical Sciences.* **24**(12), pp.651-655.

GaStat. 2022. *General Authority for Statistics GaStat.* [Online]. [Accessed 18/7/2023]. Available from: https://www.my.gov.sa/wps/portal/snp/agencies/agencyDetails/AC138

Gauthier, G., St-Onge, C. and Tavares, W. 2016. Rater cognition: review and integration of research findings. *Medical education.* **50**(5), pp.511-522.

Gilani, S., Pankhania, K., Aruketty, M., Naeem, F., Alkhayyat, A., Akhtar, U., Chaudhary, M. and Sinha, S. 2021. Twelve tips to organise a mock OSCE. *Medical Teacher.* **4**(1), pp.1-20.

Gingerich, A., Kogan, J., Yeates, P., Govaerts, M. and Holmboe, E. 2014. Seeing the 'black box'differently: assessor cognition from three research perspectives. *Medical education.* **48**(11), pp.1055-1068.

Gleeson, F. 1997. AMEE Medical Education Guide No. 9. Assessment of clinical competence using the Objective Structured Long Examination Record (OSLER). *Medical Teacher.* **19**(1), pp.7-14.

GMC. 2023a. *Clinical and professional skills assessment (CPSA).* [Online]. Available from: https://www.gmc-uk.org/education/medical-licensing-assessment/uk-medical-schools-guide-to-the-mla/clinical-and-professional-skills-assessment-cpsa

GMC. 2023b. *General Medical Council: Professional and Linguistic Assessments Board (PLAB).* [Online]. [Accessed 19 July 2023]. Available from: https://www.gmc-uk.org/registration-and-licensing/join-the-register/plab

Goh, H.S., Ng, E., Tang, M.L., Zhang, H. and Liaw, S.Y. 2022. Psychometric testing and cost of a five-station OSCE for newly graduated nurses. *Nurse Education Today.* **112**, p105326.

Gormley, G. 2011. Summative OSCEs in undergraduate medical education. *The Ulster medical journal.* **80**(3), p127.

Gorsky, M. and Mold, A. 2020. Documentary Analysis. In: Pope, C. and Mays, N. eds. *Qualitative Research in Health Care.* 4th ed. Chichester: Wiley.

Gottlie, M., Haas, M.R.C., Daniel, M. and Chan, T.M. 2021. The scoping review: A flexible, inclusive, and iterative approach to knowledge synthesis. **5**(3), pp.1-6.

Gough, D., Oliver, S. and Thomas, J. 2017. *An Introduction to Systematic Reviews.* 2nd ed. California: SAGE Publications.

Graham, I.D., Beardall, S., Carter, A.O., Tetroe, J. and Davies, B. 2003. The state of the science and art of practice guidelines development, dissemination and evaluation in Canada. *Journal of evaluation in clinical practice.* **9**(2), pp.195-202.

Graham, R., Zubiaurre Bitzer, L.A. and Anderson, O.R. 2013. Reliability and predictive validity of a comprehensive preclinical OSCE in dental education. *Journal of dental education.* **77**(2), pp.161-167.

Greiner, L.E. 1972. *Evolution and Revolution as Organizations Grow.* United States: Harvard Business Review.

Griesser, M.J., Beran, M.C., Flanigan, D.C., Quackenbush, M., Van Hoff, C. and Bishop, J.Y. 2012. Implementation of an Objective Structured Clinical Exam (OSCE) into Orthopedic Surgery Residency Training. *Journal of Surgical Education.* **69**(2), pp.180-189.

Grimm, P. 2010. Social desirability bias. *Wiley international encyclopedia of marketing.*

Groene, O.R., Bergelt, C. and Ehrhardt, M. 2022. How good are medical students at communicating risk? An implementation study at three German medical schools. *Patient Education and Counseling.* **105**(1), pp.128-135.

Guest, G., MacQueen, K.M. and Namey, E.E. 2012. *Applied Thematic Analysis.* London: SAGE Publications.

Gupta, P., Dewan, P. and Singh, T. 2010. Objective structured clinical examination (OSCE) revisited. *Indian pediatrics.* **47**(11), pp.911-920.

Haas, M.R., Munzer, B.W., Santen, S.A., Hopson, L.R., Haas, N.L., Overbeek, D., Peterson, W.J., Cranford, J.A. and Huang, R.D. 2020. DidacticsRevolution: Applying Kotter's 8-Step Change Management Model to Residency Didactics. *Western Journal of Emergency Medicine.* **21**(1), p65.

Hadi, M.A., Ali, M., Haseeb, A., Mohamed, M.M., Elrggal, M.E. and Cheema, E. 2018. Impact of test anxiety on pharmacy students' performance in Objective Structured Clinical Examination: a cross-sectional survey. *International Journal of Pharmacy Practice.* **26**(2), pp.191-194.

Hamdy, H., Telmesani, A.W., Al Wardy, N., Abdel-Khalek, N., Carruthers, G., Hassan, F., Kassab, S., Abu-Hijleh, M., Al-Roomi, K., O'malley, K., El Din Ahmed, M.G., Raj, G.A., Rao, G.M. and Sheikh, K. 2010. Undergraduate medical education in the Gulf Cooperation Council: A multi-countries study (Part 1). *Medical Teacher.* **32**(3), pp.219-224.

Hamron, D.A.b.G.b.S. 2018. Requirements for the autonomy of emerging Saudi universities from the point of view of professors of administration and educational leadership in Saudi universities. *Journal of Educational Sciences. at King Saud University.* **30**(453 - 478).

Hanson, W.E., Creswell, J.W., Clark, V.L.P., Petska, K.S. and Creswell, J.D. 2005. Mixed Methods Research Design in Counseling Psychology. *Journal of Counseling Psychology.* **25**(2), pp.224-235.

Harden. 1988. What is an OSCE? *Medical Teacher.* **10**(1), pp.19-22.

Harden. 1990. Twelve tips for organizing an Objective Structured Clinical Examination (OSCE). *Medical Teacher.* **12**(3-4), pp.259-264.

Harden. 2015. Misconceptions and the OSCE. *Medical Teacher.* **37**(7), pp.608-610.

Harden. 2016. Revisiting 'Assessment of clinical competence using an objective structured clinical examination (OSCE)'. *Medical Education.* **50**(4), pp.376-379.

Harden, Lilley, P. and Patricio, M. 2015. *The Definitive Guide to the OSCE: The Objective Structured Clinical Examination as a performance assessment.* 1 ed. Edinburgh: Elsevier.

Harden, Stevenson, M., Downie, W.W. and Wilson, G.M. 1975. Assessment of clinical competence using objective structured examination. *British Medical Journal.* **1**(5955), pp.447-451.

Harden, R.M. and Gleeson, F.A. 1979. Assessment of medical competence using an Objective Structure Clinical Examination (OSCE). *Medical Education.* **13**(1), pp.39-54.

Harrison, C.J., Könings, K.D., Schuwirth, L.W.T., Wass, V. and van der Vleuten, C.P.M. 2017. Changing the culture of assessment: the dominance of the summative assessment paradigm. *BMC Medical Education.* **17**(1), p73.

Harvey, L. 2004. The power of accreditation: Views of academics. *Journal of Higher Education Policy and Management.* **26**(2), pp.207-223.

Hattie, J. and Clarke, S. 2018. *Visible learning: feedback.* 1st ed. London: Routledge.

Heal, C., D'Souza, K., Banks, J., Malau-Aduli, B.S., Turner, R., Smith, J., Bray, E., Shires, L. and Wilson, I. 2019. A snapshot of current Objective Structured Clinical Examination (OSCE) practice at Australian medical schools. *Medical Teacher.* **41**(4), pp.441-447.

Hernandez-Galvez, D.C. and Roldan-Valadez, E. 2019. Mexican ENARM: performance comparison of public vs. private medical schools, geographic and socioeconomic regions. *salud pública de méxico.* **61**(5), pp.637-647.

Hidayah, R.N. 2018. *Impact of the national medical licensing examination in Indonesia: perspectives from students, teachers, and medical schools*. thesis, University of Leeds.

Higgins, J.P.T. 2019. *Cochrane Handbook for Systematic Reviews of Interventions.* NJ, USA: Wiley.

Hijazi, M. and Downing, S.M. 2008. Objective structured clinical examinations as an assessment method in residency training: Practical considerations. *Annals of Saudi Medicine.* **28**(3), pp.192-199.

Hodges, B. 2003a. OSCE! Variations on a theme by Harden. *Medical Education.* **37**(12), pp.1134-1140.

Hodges, B. 2003b. Validity and the OSCE. *Medical Teacher.* **25**(3), pp.250-255.

Hodges, B. and McIlroy, J.H. 2003. Analytic global OSCE ratings are sensitive to level of training. *Medical education.* **37**(11), pp.1012-1016.

Hodges, B., Paul, R., Ginsburg, S. and Members, T.O.C.G. 2019. Assessment of professionalism: From where have we come–to where are we going? An update from the Ottawa Consensus Group on the assessment of professionalism. *Medical Teacher.* **41**(3), pp.249-255.

Hodges, B.D., Ginsburg, S., Cruess, R., Cruess, S., Delport, R., Hafferty, F., Ho, M.-J., Holmboe, E., Holtman, M. and Ohbu, S. 2011. Assessment of professionalism: recommendations from the Ottawa 2010 Conference. *Medical teacher.* **33**(5), pp.354-363.

Holzweiss, P.C., Bustamante, R. and Fuller, M.B. 2016. Institutional Cultures of Assessment: A Qualitative Study of Administrator Perspectives. *Journal of Assessment and Institutional Effectiveness.* **6**(1), pp.1-27.

Homer, M. 2022. Pass/fail decisions and standards: the impact of differential examiner stringency on OSCE outcomes. *Advances in Health Sciences Education.* **27**(2), pp.457-473.

Homer, M. 2023. Setting defensible minimum-stations-passed standards in OSCE-type assessments. *Medical Teacher.* (ahead-of-print), pp.1-7.

Homer, M. and Darling, J.C. 2016. Setting standards in knowledge assessments: comparing Ebel and Cohen via Rasch. *Medical Teacher.* **38**(12), pp.1267-1277.

Homer, M., Fuller, R., Hallam, J. and Pell, G. 2020a. Setting defensible standards in small cohort OSCEs: understanding better when borderline regression can 'work'. *Medical teacher.* **42**(3), pp.306-315.

Homer, M., Fuller, R., Hallam, J. and Pell, G. 2020b. Shining a spotlight on scoring in the OSCE: Checklists and item weighting. *Medical Teacher.* **42**(9), pp.1037-1042.

Homer, M., Fuller, R. and Pell, G. 2018. The benefits of sequential testing: Improved diagnostic accuracy and better outcomes for failing students. *Medical teacher.* **40**(3), pp.275-284.

Homer, M., Pell, G. and Fuller, R. 2017. Problematizing the concept of the "borderline" group in performance assessments. *Medical Teacher.* **39**(5), pp.469-475.

Homer, M., Pell, G., Fuller, R. and Patterson, J. 2016. Quantifying error in OSCE standard setting for varying cohort sizes: A resampling approach to measuring assessment quality. *Medical Teacher.* **38**(2), pp.181-188.

Horita, S., Park, Y.-S., Son, D. and Eto, M. 2021. Computer-based test (CBT) and OSCE scores predict residency matching and National Board assessment results in Japan. *BMC Medical Education.* **21**(1), pp.1-8.

Huang, Y.-S., Liu, M., Huang, C.-H. and Liu, K.-M. 2007. Implementation of an OSCE at Kaohsiung Medical University. *The Kaohsiung Journal of Medical Sciences.* **23**(4), pp.161-169.

Humphrey-Murto, S., Touche, C. and Smee, S. 2013. Objective structured clinical examination. In: Walsh, K. ed. *Oxford Textbook of Medical Education.* 1st ed. Oxford: Oxford University Press.

Hussain, S., Shen, L., Akram, T., Haider, M., Hussain, S. and Ali, M. 2016. Kurt Lewin's process model for organizational change: The role of leadership and employee involvement: A critical review. *Journal of Innovation & Knowledge.* **3**(3), pp.123-127.

Huwendiek, S., Jung, D., Schirlo, C., Huber, P., Balestra, G., Guttormsen, S. and Berendonk, C. 2020. The introduction of a standardised national licensing exam as a driver of change in medical education: A qualitative study from Switzerland. *Medical teacher.* **42**(10), pp.1163-1170.

Hyde, S., Fessey, C., Boursicot, K., MacKenzie, R. and McGrath, D. 2022. OSCE rater cognition–an international multi-centre qualitative study. *BMC Medical Education.* **22**, pp.1-10.

Ibrahim, Banjar, S., Al-Ghamdi, A., Al-Darmasi, M., Khoja, A., Turkistani, J., Arif, R., Al-Sebyani, A., Musawa, A.A. and Basfar, W. 2014. Medical students preference of problem-based learning or traditional lectures in King Abdulaziz University, Jeddah, Saudi Arabia. *Ann Saudi Med.* **34**(2), pp.128-133.

Ibrahim, N.K., Al-Sharabi, B.M., Al-Asiri, R.A., Alotaibi, N.A., Al-Husaini, W.I., Al-Khajah, H.A., Rakkah, R.M. and Turkistani, A.M. 2015. Perceptions of clinical years' medical students and interns towards assessment methods used in King Abdulaziz University, Jeddah. *Pakistan Journal of Medical Sciences.* **31**(4), pp.757-762.

Ilgen, J.S., Ma, I.W., Hatala, R. and Cook, D.A.J.M.e. 2015. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. **49**(2), pp.161-173.

Inayah, A.T., Anwer, L.A., Shareef, M.A., Nurhussen, A., Alkabbani, H.M., Alzahrani, A.A., Obad, A.S., Zafar, M. and Afsar, N.A. 2017. Objectivity in subjectivity: do students' self and peer assessments correlate with examiners' subjective and objective assessment in clinical skills? A prospective study. *BMJ Open.* **7**(5), pp.1-7.

Iqbal, M., Khizar, B. and Zaidi, Z. 2009. Revising an objective structured clinical examination in a resource-limited Pakistani Medical School. *Education for Health.* **22**(1), p209.

Ismail. 2001. Views of undergraduate students on objective structured clinical examination in neurology: a preliminary report. *J Family Community Med.* **8**(3), pp.79-82.

Jafarzadeh, A. 2009. Designing the OSCE method for evaluation of practical immunology course of medical students: In comparison to written-MCQ and oral examination. *Rawal Medical Journal.* **34**(2), pp.219-222.

Jiang, Z., Shi, D. and Distefano, C. 2021. A Short Note on Optimizing Cost-Generalizability via a Machine-Learning Approach. *Educational and Psychological Measurement.* **81**(6), pp.1221-1233.

Johnson, J.L., Adkins, D. and Chauvin, S. 2020. A review of the quality indicators of rigor in qualitative research. *American journal of pharmaceutical education.* **84**(1), pp.138-146.

Johnson, R.L., Penny, J.A. and Gordon, B. 2008. *Assessing performance: Designing, scoring, and validating performance tasks.* New York: Guilford Press.

Jolly, B. 2014. Written Assessment. In: Swanwick, T. ed. *Understanding Medical Education: Evidence, Theory and Practice.* 2nd ed. West Sussex: Wiley Blackwell, pp.255-277.

Jorm, C. and Roberts, C. 2018. Using complexity theory to guide medical school evaluations. *Academic Medicine.* **93**(3), pp.399-405.

Joy, S., Ravindran, R.M., Vijayan, D. and Ugargol, A.P. 2007. Mushrooming of private medical schools in India, the present student profile and cost of medical education and its impact on equity, health care accessibility, cost and perceived quality of health care Health Care Accessibility, Cost and Perceived Quality of Health Care. In: *iHEA 2007 6th World Congress: Explorations in Health Economics Paper*.

Kane, M. 1992. An argument-based approach to validity. *Psychological bulletin.* **112**(3), pp.527-535.

Kane, M. 2006. Validation. In: Brennan, R.L. ed. *Educational Measurement.* 4th ed. Westport, CT: American Council on Education and Praeger, pp.17-64.

Kane, M. 2010. Validity and fairness. *Language Testing.* **27**(2), pp.177-182.

Kang, S.P., Chen, Y., Svihla, V., Gallup, A., Ferris, K. and Datye, A.K. 2022. Guiding change in higher education: An emergent, iterative application of Kotter's change model. *Studies in Higher Education.* **47**(2), pp.270-289.

Karam, Y., Park, Y.S., Tekian, A. and Youssef, N. 2018. Evaluating the validity evidence of an OSCE: Results from a new medical school. *BMC Medical Education.* **18**(1), pp.313-320.

Kearney, G.P., Johnston, J.L., Hart, N.D., Corman, M.K. and Gormley, G.J. 2018. Protocol: exploring the objective structured clinical examination (OSCE) using institutional ethnography. *International Journal of Educational Research.* **88**, pp.42-47.

Khairy, G.A. 2004. Feasibility and acceptability of objective structured clinical examination (osce) for a large number of candidates: experience at a university hospital. *Journal of family & community medicine.* **11**(2), pp.75-78.

Khan, K.Z., Gaunt, K., Ramachandran, S. and Pushkar, P. 2013a. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: Organisation & Administration. *Medical Teacher.* **35**(9), pp.e1447-e1463.

Khan, K.Z., Ramachandran, S., Gaunt, K. and Pushkar, P. 2013b. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: An historical and theoretical perspective. *Medical Teacher.* **35**(9), pp.e1437-e1446.

Khapre, M.P., Sabane, H., Singh, S., Katyal, R., Kapoor, A. and Badyal, D.K. 2020. Faculty's perspective on skill assessment in undergraduate medical education: Qualitative online forum study. *Journal of Education and Health Promotion.* **9**(1), pp.20-27.

Kiger, M.E. and Varpio, L. 2020. Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical Teacher.* **42**(8), pp.1-9.

King, N. and Brooks, J. 2018. Thematic Analysis in Organisational Research. In: Cassell, C., et al. eds. *The SAGE Handbook of Qualitative Business and Management Research Methods.* London: SAGE Publications.

King, N., Horrocks, C. and Brooks, J. 2018. *Interviews in Qualitative Research.* 2nd ed. Cambridge, UK: SAGE Publications.

Kirchherr, J. 2019. *The Lean PhD.* 1st ed. London, UK: Palgrave.

Kitto, S., Alexanian, J. and Goldman, J. 2022. Attuning to the social world: ethnography in health professions education research. In: Cleland, J. and Durning, S.J. eds. *Researching Medical Education.* 2nd ed. Chichester: Wiley.

Kitto, S.C., Chesters, J. and Grbich, C. 2008. Quality in qualitative research. *Medical journal of Australia.* **188**(4), pp.243-246.

Komasawa, N., Terasaki, F., Nakano, T. and Kawata, R. 2020. Relationships between objective structured clinical examination, computer-based testing, and clinical clerkship performance in Japanese medical students. *PLoS One.* **15**(3), pp.1-8.

Kotter, J.P. 1995. Leading change: Why transformation efforts fail. *Harvard business review.* **73**(2), pp.59-67.

Kotter, J.P. 2012. *Leading Change.* Boston: Harvard Business Review Press.

Kramer, A., Muijtjens, A., Jansen, K., Düsman, H., Tan, L. and Van Der Vleuten, C. 2003. Comparison of a rational and an empirical standard setting procedure for an OSCE. *Medical Education.* **37**(2), pp.132-139.

Krueger, R.A. and Casey, M.A. 2015. *Focus groups : a practical guide for applied research.* 5th ed. Thousand Oaks, California: SAGE Publications.

Kua, J., Teo, W. and Lim, W.S. 2022. Learning experiences of adaptive experts: a reflexive thematic analysis. *Advances in Health Sciences Education.* **27**(5), pp.1345-1359.

Kübler-Ross, E. 1969. *On Death and Dying: What the Dying Have to Teach Doctors, Nurses, Clergy, and Their Own Families.* United Kingdom: Touchstone, Simon & Schuster.

Kunnan, A.J. 2000. Fairness and justice for all. In: Kunnan, A.J. ed. *Fairness and Validation in Language Assessment.* Cambridge, UK: Cambridge University Press, pp.1-14.

Kuper, A., Lingard, L. and Levinson, W. 2008a. Critically appraising qualitative research. *Bmj.* **337**, pp.687-692.

Kuper, A., Reeves, S. and Levinson, W. 2008b. An introduction to reading and appraising qualitative research. *Bmj.* **337**(7666), pp.404-407.

Kvale, S. 2018. *Doing Interviews.* 2nd ed. London, UK: SAGE Publications.

Lane, S., Raymond, M.R. and Haladyna, T.M. 2016. *Handbook of test development.* 2nd ed. New York: Routledge.

Latjatih, N.H.F., Roslan, N.S., Jahn Kassim, P.S. and Adam, S.K. 2022. Medical Students' Perception and Satisfaction on Peer-Assisted Learning in Formative OSCE and Its Effectiveness in Improving Clinical Competencies. *Journal of Applied Research in Higher Education.* **14**(1), pp.171-179.

Leavy, P. 2017. *Research Design: Quantitative, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approaches.* 1st ed. London, UK: The Guilford Press.

Lee, K., Liu, W., Yen, J., Liu, C., Liu, K. and Lai, C. 2008. The experience of an objective, structured clinical examination at Kaohsiung Medical University. *Kaohsiung Journal of Medical Sciences.* **24**(12), pp.624-626.

Lee, M.C.Y. 2016. Finding cultural harmony in interviewing: the wisdom of the middle way. *International Journal of Research & Method in Education.* **39**(1), pp.38-57.

Levac, D., Colquhoun, H. and O'Brien, K. 2010. Scoping studies: advancing the methodology. **5**(1), p69.

Lewin, K. 1947. Group decision and social change. *T. Newcomb and F. Hartley, Readings in Social Psychology. New York: Henry Holt and Company.*

Lincoln, Y.S. and Guba, E.G. 1985. *Naturalistic inquiry.* London: SAGE.

Ma, I.W., Zalunardo, N., Pachev, G., Beran, T., Brown, M., Hatala, R. and McLaughlin, K. 2012. Comparing the use of global rating scale with checklists for the assessment of central venous catheterization skills using simulation. *Advances in health sciences education.* **17**(4), pp.457-470.

MacLean, S., Geddes, F., Kelly, M. and Della, P. 2018. Simulated patient training: Using inter-rater reliability to evaluate simulated patient consistency in nursing education. *Nurse education today.* **62**, pp.85-90.

Maggio, L.A., Larsen, K., Thomas, A., Costello, J.A. and Artino Jr, A.R. 2020. Scoping reviews in medical education: A scoping review. *Medical education.* **55**(6), pp.689-700.

Mahmood, T. 2018. What models of change can be used to implement change in postgraduate medical education? *Advances in Medical Education and Practice.* **9**, pp.175-178.

Majumder, M.A.A., Kumar, A., Krishnamurthy, K., Ojeh, N., Adams, O.P. and Sa, B. 2019. An evaluative study of objective structured clinical examination (OSCE): students and examiners perspectives. *Advances in medical education and practice.* **10**, pp.387-397.

Malau-Aduli, B.S., Hays, R.B., D'Souza, K., Saad, S.L., Rienits, H., Celenza, A. and Murphy, R. 2023. Twelve tips for improving the quality of assessor judgements in senior medical student clinical assessments. *Medical Teacher.* (ahead-of-print), pp.1-5.

Malau-Aduli, B.S., Hays, R.B., D'Souza, K., Smith, A.M., Jones, K., Turner, R., Shires, L., Smith, J., Saad, S. and Richmond, C. 2021. Examiners' decision-making processes in observation-based clinical examinations. *Medical Education.* **55**(3), pp.344-353.

Mamatha, S.D., Rajalakshmi, R. and Ramya, C.M. 2017. First year MBBS student's perception of objective structured clinical examination and objective structured practical examination. *Indian journal of physiology and pharmacology.* **61**(4), pp.440-444.

Manchester, J., Gray-Miceli, D.L., Metcalf, J.A., Paolini, C.A., Napier, A.H., Coogle, C.L. and Owens, M.G. 2014. Facilitating Lewin's change model with collaborative evaluation in promoting evidence based practices of health professionals. *Evaluation and program planning.* **47**, pp.82-90.

Mann, C., Canny, B.J., Reser, D.H. and Rajan, R. 2013. Poorer verbal working memory for a second language selectively impacts academic achievement in university medical students. *PeerJ.* **12**(1), pe22.

Marshall, C. and Rossman, G.B. 2014. *Designing Qualitative Research.* 6th ed. London: SAGE.

MCC. 2023. *The MCC ceases delivery of the MCCQE Part II | Medical Council of Canada.* [Online]. [Accessed 16 May]. Available from: https://mcc.ca/news/mcc-ceases-delivery-of-the-mccqe-part-ii/

McCowan, T. 2004. The growth of private higher education in Brazil: implications for equity and quality. *Journal of Education Policy.* **19**(4), pp.453-472.

McKinley, D.W. and Norcini, J.J. 2014. How to set standards on performance-based examinations: AMEE Guide No. 85. *Medical teacher.* **36**(2), pp.97-110.

McManus, I.C., Thompson, M. and Mollon, J. 2006. Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP (UK) clinical examination (PACES) using multi-facet Rasch modelling. *BMC Medical Education.* **6**(1), pp.1-22.

Merriam-Webster. 2019. *Definition of DESIGNER.* [Online]. Available from: https://www.merriam-webster.com/dictionary/designer

Messick, S. 1989. VALIDITY. In: Linn, R.L. ed. *Educational Measurement.* 3rd ed. New York: American Council on Education and Macmillan, pp.13-103.

Mihas, P. 2019. *Learn to build a codebook for a generic qualitative study.* [Online]. London: SAGE Publications.

Miles, M.B., Huberman, A.M. and Saldana, J. 2018. *Qualitative Data Analysis A Methods Sourcebook.* 4th ed. London: SAGE Publications, Inc.

Milic, J.V., Ehrler, B., Molina, C., Saliba, M. and Bisquert, J. 2020. Online meetings in times of global crisis: Toward sustainable conferencing. *ACS Energy Letters.* **5**(6), pp.2024–2026.

Miller, G.E. 1990. The assessment of clinical skills/competence/performance

Academic Medicine. **65**(9), pp.S63–S67.

Mirza, D. and Hashim, M. 2010. Communication skills training in English alone can leave Arab medical students unconfident with patient communication in their native language. *Education for Health.* **23**(2), p450.

Miyata, H. and Kai, I. 2009. Reconsidering evaluation criteria for scientific adequacy in health care research: An integrative framework of quantitative and qualitative criteria. *International Journal of Qualitative Methods.* **8**(1), pp.64-75.

MOE. 2022. *Ministry of Education, Kingdom of Saudi Arabia* [Online]. [Accessed 29 July]. Available from: https://www.moe.gov.sa/en/Pages/default.aspx

MOE. 2023. *Ministry of Education, Saudi Arabia.* [Online]. [Accessed 24 July]. Available from: https://www.moe.gov.sa/en/Pages/default.aspx

Moeen-uz-zafar, Shammari, O.A. and Aljarallah, B. 2015. Evaluation of Interactive OSCE for Medical Students in the Subject of Medicine; Reliability and Validity in the Setting of Internal vs. External Examiners. *Sci Med Central.* **2**(4), p1030.

MOFA. 2023. *KSA History.* [Online]. Available from: https://www.mofa.gov.sa/en/ksa/Pages/history.aspx

Mohamed, M. 2003. Medical Schools: The supply and availability of qualified human resources-challenges and opportunities. *The Malaysian Journal of Medical Sciences: MJMS.* **10**(1), pp.1-3.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G. and PRISMA Group*, t. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine.* **151**(4), pp.264-269.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A. and Group, P.-P. 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews.* **4**(1), pp.1-9.

Moltner, A., Lehmann, M., Wachter, C., Kurczyk, S., Schwill, S. and Loukanova, S. 2020. Formative assessment of practical skills with peer-assessors: quality features of an OSCE in general medicine at the Heidelberg Medical Faculty. *GMS Journal for Medical Education.* **37**(4), pDoc42.

Mondal, R., Sarkar, S., Nandi, M. and Hazra, A. 2012. Comparative Analysis Between Objective Structured Clinical Examination (OSCE) and Conventional Examination (CE) As a Formative Evaluation Tool in Pediatrics in Semester Examination for Final MBBS Students. *Kathmandu University Medical Journal.* **10**(1), pp.53-56.

Moriguchi, Y. 2022. Beyond bias to Western participants, authors, and editors in developmental science. *Infant and Child Development.* **31**(1), pe2256.

Mufti, M.H. 2000. *Healthcare Development Strategies in the Kingdom of Saudi Arabia.* New York: Springer US.

Muijs, D. and Rumyantseva, N. 2014. Coopetition in education: Collaborating in a competitive environment. *Journal of educational Change.* **15**, pp.1-18.

Mulimani, P. 2019. Publication bias towards Western populations harms humanity. *Nature Human Behaviour.* **3**(10), pp.1026-1027.

Nadin, S. and Cassell, C. 2004. Using data matrices. In: Cassell, C. and Symon, G. eds. *Essential Guide to Qualitative Methods in Organizational Research.* London: SAGE.

NCAAA. 2023a. *National Center for Academic Accreditation and Evaluation.* [Online]. [Accessed 24 June]. Available from: https://etec.gov.sa/ncaaa

NCAAA. 2023b. *NCAAA Center: Results for programmatic accreditation for MBBS programmes.* [Online]. [Accessed 18 July 2023]. Available from: https://www.etec.gov.sa/ncaaa/accreditationsresults

Ndoye, A. and Parker, M.A. 2010. Creating and sustaining a culture of assessment. *Planning for Higher Education.* **38**(2), p28.

Nelson, A. 1976. Candidates need 'workable' theories. *The Boston Herald*.

Nestel, D., Kneebone, R., Nolan, C., Akhtar, K. and Darzi, A. 2011. Formative assessment of procedural skills: students' responses to the objective structured clinical examination and the integrated performance procedural instrument. *Assessment & Evaluation in Higher Education.* **36**(2), pp.171-183.

Newble, D. 1988. Eight years' experience with a structured clinical examination. *Medical Education.* **22**(3), pp.200-204.

Newble, D. 1998. Assessment. In: Jolly, B. and Rees, L.H. eds. *Medical education in the millennium.* Oxford University Press, USA, pp.131–142.

Newble, D. 2004. Techniques for measuring clinical competence: Objective structured clinical examinations. *Medical Education.* **38**(2), pp.199-203.

Newble, D. and Swanson, D. 1988. Psychometric characteristics of the objective structured clinical examination. *Medical education.* **22**(4), pp.325-334.

Ngim, C.F., Fullerton, P.D., Ratnasingam, V., Arasoo, V.J.T., Dominic, N.A., Niap, C.P.S. and Thurairajasingam, S. 2021. Feedback after OSCE: A comparison of face to face versus an enhanced written feedback. *BMC medical education.* **21**, pp.1-9.

Norcini, J., Anderson, B., Bollela, V., Burch, V., Costa, M.J., Duvivier, R., Galbraith, R., Hays, R., Kent, A., Perrott, V. and Roberts, T. 2011. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher.* **33**(3), pp.206-214.

Norcini, J., Anderson, M.B., Bollela, V., Burch, V., Costa, M.J., Duvivier, R., Hays, R., Palacios Mackay, M.F., Roberts, T. and Swanson, D. 2018. 2018 Consensus framework for good assessment. *Medical Teacher.* **40**(11), pp.1102-1109.

Norcini, J. and Ben-David, M.F. 2021. Concepts in assessment. In: Dent, J., et al. eds. *A practical guide for medical teachers.* 6th ed. London: Elsevier Health Sciences.

Norcini, J.J. 2003. Setting standards on educational tests. *Medical Education.* **37**(5), pp.464-469.

Norman, G. 2002. Research in medical education: three decades of progress. **324**(7353), pp.1560-1562.

Norman, G. and Eva, K.W. 2018. Quantitative research methods in medical education. *Understanding medical education: evidence, theory, and practice.* pp.405-425.

Nowell, L.S., Norris, J.M., White, D.E. and Moules, N.J. 2017. Thematic analysis: Striving to meet the trustworthiness criteria. *International journal of qualitative methods.* **16**(1), pp.1-13.

NQF. 2020. *National Qualification Framework.* [Online]. [Accessed 24 June]. Available from: https://spapi.etec.gov.sa:2443/en/MediaAssets/NCAAA/General%20Documents/NQF-KSA%202020.pdf?csf=1&e=UgktO8

Nyangeni, T., van Rooyen, D.R., Du Plessis, A.H. and ten Ham-Baloyi, W. 2022. Quality management of objective structured clinical examinations: An integrative literature review. *Nurse Education Today.* **109**, p105223.

O'Leary, Z. 2017. *Am I The Only One Struggling To Write A Literature Review?* Available from: https://methods.sagepub.com/video/am-i-the-only-one-struggling-to-write-a-literature-review

O'Leary, Z. 2021. *The Essential Guide to Doing Your Research Project.* 4th ed. London: SAGE Publications.

Obeidat, A.S., Alhaqwi, A.I. and Abdulghani, H.M. 2015. Reprioritizing current research trends in medical education: A reflection on research activities in Saudi Arabia. *Medical Teacher.* **37**(1), pp.S5-S8.

Odiaga, J., Guglielmo, M.J., Catrambone, C., Gierlowski, T., Bruti, C., Richter, L. and Miller, J. 2021. KOTTER'S CHANGE MODEL IN HIGHER EDUCATION: TRANSFORMING SILOED EDUCATION TO A CULTURE OF INTERPROFESSIONALISM. *Journal of Organizational Culture, Communications and Conflict.* **25**(2), pp.1-7.

Olafsen, A.H., Nilsen, E.R., Smedsrud, S. and Kamaric, D. 2021. Sustainable development through commitment to organizational change: The implications of organizational culture and individual readiness for change. *Journal of Workplace Learning.* **33**(3), pp.180-196.

Olmos-Vega, F.M., Stalmeijer, R.E., Varpio, L. and Kahlke, R. 2022. A practical guide to reflexivity in qualitative research: AMEE Guide No. 149. *Medical Teacher.* **45**(3), pp.241-251.

Onwudiegwu, U. 2018. OSCE: DESIGN, DEVELOPMENT AND DEPLOYMENT. *J West Afr Coll Surg.* **8**(1), pp.1-22.

Orovec, A., Bishop, A., Scott, S.A., Wilson, D., Richardson, C.G., Oxner, W. and Glennie, R.A. 2022. Validation of a Surgical Objective Structured Clinical Examination (S-OSCE) Using Convergent, Divergent, And Trainee-Based Assessments of Fidelity. *Journal of Surgical Education.* **79**(4), pp.1000-1008.

Ossenberg, C., Henderson, A. and Mitchell, M. 2019. What attributes guide best practice for effective feedback? A scoping review. *Advances in Health Sciences Education.* **24**(2), pp.383-401.

Othman, Z. and Hamid, F.Z.A. 2018. Dealing with un (expected) ethical dilemma: Experience from the field. *The Qualitative Report.* **23**(4), pp.733-741.

Parikh, A., McReelis, K. and Hodges, B. 2001. Student feedback in problem based learning: a survey of 103 final year students across five Ontario medical schools. *Medical education.* **35**(7), pp.632-636.

Parish, S.J., Ramaswamy, M., Stein, M.R., Kachur, E.K. and Arnsten, J.H. 2006. Teaching about substance abuse with objective structured clinical exams. *Journal of general internal medicine.* **21**, pp.453-459.

Park, S., Griffin, A. and Gill, D. 2012. Working with words: exploring textual analysis in medical education research. *Medical Education.* **46**(4), pp.372-380.

Park, S.Y., Lee, S.-H., Kim, M.-J., Ji, K.-H. and Ryu, J.H. 2021. Comparing the cut score for the borderline group method and borderline regression method with norm-referenced standard setting in an objective structured clinical examination

in medical school in Korea. *Journal of Educational Evaluation for Health Professions.* **18**, pp.25-35.

Patricio, Juliao, M., Fareleira, F. and Carneiro, A.V. 2013. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? . *Medical Teacher.* **35**, pp.503-514.

Patricio, M., Julião, M., Fareleira, F., Young, M., Norman, G. and Vaz Carneiro, A. 2009. A comprehensive checklist for reporting the use of OSCEs. *Medical Teacher.* **31**(2), pp.112-124.

Patricio, M.F. 2012. *A Best Evidence Medical Education (BEME) Systematic Review on the feasibility, reliability and validity of the Objective Structured Clinical Examination (OSCE) in undergraduate medical studies*. thesis, Universidade de Lisboa.

Patton, M.Q. 2015. *Qualitative evaluation and research methods: Integrating Theory and Practice.* 4th ed. California SAGE Publications.

Pell, G., Fuller, R., Homer, M. and Roberts, T. 2010. How to measure the quality of the OSCE: A review of metrics: AMEE guide no. 49. *Medical Teacher.* **32**(10), pp.802-811.

Pell, G., Fuller, R., Homer, M. and Roberts, T. 2013. Advancing the objective structured clinical examination: Sequential testing in theory and practice. *Medical Education.* **47**(6), pp.569-577.

Pell, G., Homer, M. and Roberts, T. 2008. Assessor training: its effects on criterion-based assessment in a medical context. *International Journal of Research & Method in Education.* **31**(2), pp.143-154.

Pell, G. and Roberts, T.E. 2006. Setting standards for student assessment. *International Journal of Research & Method in Education.* **29**(1), pp.91-103.

Peters, M.D.J., Godfrey, C.M., Khalil, H., McInerney, P., Parker, D. and Soares, C.B. 2015. Guidance for conducting systematic scoping reviews. **13**(3), pp.141-146.

Peters, T.J. and Waterman, R.H. 1982. *In Search of Excellence: Lessons from America's Best-Run Companies.* New York: HarperCollins.

Pierre, R.B., Wierenga, A., Barton, M., Branday, J.M. and Christie, C.D.C. 2004. Student evaluation of an OSCE in paediatrics at the University of the West Indies, Jamaica. *BMC Medical Education.* **4**(1), p22.

Poenaru, D., Morales, D., Richards, A. and O'Connor, H.M. 1997. Running an objective structured clinical examination on a shoestring budget. *American Journal of Surgery.* **173**(6), pp.538-541.

Ponnamperuma, G.G., Karunathilake, I.M., McAleer, S. and Davis, M.H. 2009. The long case and its modifications: a literature review. **43**(10), pp.936-941.

Price, T., Lynn, N., Coombes, L., Roberts, M., Gale, T., de Bere, S.R. and Archer, J. 2018. The international landscape of medical licensing examinations: a typology derived from a systematic review. *International journal of health policy and management.* **7**(9), pp.782-790.

PubMed. 2023. *Journals indexed in the National Library of Medicine (NLM) Catalogue* [Online]. [Accessed 4 July]. Available from: https://www.ncbi.nlm.nih.gov/nlmcatalog

Pugh, D., Desjardins, I. and Eva, K. 2018. How do formative objective structured clinical examinations drive learning? Analysis of residents' perceptions. *Medical Teacher.* **40**(1), pp.45-52.

Pugh, D. and Smee, S. 2013. *Guidelines for the development of objective structured clinical examination (OSCE) cases.* [Online]. Canada: Medical Council of Canada. [Accessed 31 July 2023]. Available from: https://shorturl.at/jruFT

Raheel, H. and Naeem, N. 2013. Assessing the Objective Structured Clinical Examination: Saudi family medicine undergraduate medical students' perceptions of the tool. *Journal of the Pakistan Medical Association.* **63**(10), pp.1281-1284.

Ravinetto, R. and Singh, J.A. 2022. Responsible dissemination of health and medical research: some guidance points. *BMJ Evidence-Based Medicine.* **28**(3), pp.144-147.

Raymond, M.R. and Grande, J.P. 2019. A practical guide to test blueprinting. *Medical teacher.* **41**(8), pp.854-861.

Regehr, G., MacRae, H., Reznick, R.K. and Szalay, D. 1998. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic Medicine.* **73**(9), pp.993-997.

Rehman, R., Syed, S., Iqbal, A. and Rehan, R.R. 2012. Perception and performance of medical students in Objective structured practical examination and viva voce. *Pakistan Journal of Physiology.* **8**(2), pp.33-36.

Reid, K., Smallwood, D., Collins, M., Sutherland, R. and Dodds, A. 2016. Taking OSCE examiner training on the road: reaching the masses. *Medical education online.* **21**(1), p32389.

Rekany, A. and Aldabbagh, S. 2010. Validity and Reliability of OSCE in Evaluating Practical Performance Skills of Interns in Emergency Medicine. *Duhok Med J.* **4**(2), pp.21-29.

Rezigalla, A. 2016. Fixed pass mark: Time for change. *Sudan Medical Monitor.* **11**(4), pp.119-120.

Rhydderch, M., Elwyn, G., Marshall, M. and Grol, R. 2004. Organisational change theory and the use of indicators in general practice. *BMJ Quality & Safety.* **13**(3), pp.213-217.

Roberts, C., Newble, D., Jolly, B., Reed, M. and Hampton, K. 2006. Assuring the quality of high-stakes undergraduate assessments of clinical competence. *Medical teacher.* **28**(6), pp.535-543.

Roberts, K., Dowell, A. and Nie, J.-B. 2019. Attempting rigour and replicability in thematic analysis of qualitative research data; a case study of codebook development. *BMC Medical Research Methodology.* **19**(1), pp.66-74.

Rubin, H.J. and Rubin, I.S. 2012. *Qualitative Interviewing: The Art of Hearing Data.* 3ed ed. London: SAGE.

Rudland, J., Wilkinson, T., Smith-Han, K. and Thompson-Fawcett, M. 2008. “You can do it late at night or in the morning. You can do it at home, I did it with my flatmate.” The educational impact of an OSCE. *Medical teacher.* **30**(2), pp.206-211.

Rushforth, H.E. 2007. Objective structured clinical examination (OSCE): Review of literature and implications for nursing education. *Nurse Education Today.* **27**(5), pp.481-490.

Saeed, N., Jaffery, T. and Quadri, K.H. 2012. Exploring the process of final year objective structured clinical examination for improving the quality of assessment. *JPMA. The Journal of the Pakistan Medical Association.* **62**(8), pp.797-802.

Sainterant, O., Clisbee, M. and Julceus, E.F. 2019. Introducing the Objective Structured Clinical Examination in Haiti. *Journal of Graduate Medical Education.* **11**(4s), pp.199-200.

Saldana, J. 2021. *The Coding Manual for Qualitative Researchers.* 4th ed. London: SAGE Publications.

Sallam, A., Atwa, H., Abdelaziz, A. and Nasser, A.A. 2022. Medical Students' Perception of a Newly Implemented Objective Structured Clinical Examination (OSCE) in Orthopedic Surgery and Trauma: A Mixed-Method Study. *Journal of Ecophysiology and Occupational Health.* **22(1)**, pp.29-38.

Sanyal, B.C. and Martin, M. 2007. Quality assurance and the role of accreditation: An overview. *Report: Higher Education in the World 2007: Accreditation for Quality Assurance: What is at Stake?* [Online]. pp.3-17.

Sanyal, S. and Hisam, M.W. 2018. The impact of teamwork on work performance of employees: A study of faculty members in Dhofar University. *IOSR Journal of Business and Management.* **20**(3), pp.15-22.

SCFHS. 2023. *Saudi Commission For Health Specialties.* [Online]. [Accessed 31 Julay 2023]. Available from: https://scfhs.org.sa/sites/default/files/2022-07/SMLE%202022%20-%20Guide_0.pdf

Schleicher, I., Leitner, K., Juenger, J., Moeltner, A., Ruesseler, M., Bender, B., Sterz, J., Schuettler, K.-F., Koenig, S. and Kreuder, J.G. 2017a. Does quantity ensure quality? Standardized OSCE-stations for outcome-oriented evaluation of practical skills at different medical faculties. *Annals of Anatomy.* **212**, pp.55-60.

Schleicher, I., Leitner, K., Juenger, J., Moeltner, A., Ruesseler, M., Bender, B., Sterz, J., Schuettler, K.-F., Koenig, S. and Kreuder, J.G. 2017b. Examiner effect on the objective structured clinical exam–a study at five medical schools. *BMC medical education.* **17**(1), pp.1-7.

Schüttpelz-Brauns, K., Nühse, K., Strohmer, R. and Kaden, J.J. 2019. Training OSCE examiners: minimal effort with far-reaching results. *Medical Education.* **53**(11), pp.1153-1154.

Schuwirth and Vleuten, v.d. 2011. General overview of the theories used in assessment: AMEE Guide No. 57. *Med Teach.* **33**(10), pp.783-797.

Schuwirth, L., Norcini, J., Walker, L. and Prideaux, D. 2021. Quality Assurance of Written Assessment. In: Malau-Aduli, B.S., et al. eds. *Understanding Assessment in Medical Education through Quality Assurance.* United States: McGraw Hill LLC.

Schuwirth, L. and van der Vleuten, C. 2014. How to design a useful test: The principles of assessment. In: Swanwick, T. ed. *Understanding Medical Education: Evidence, Theory and Practice.* 2nd ed. West Sussex: Wiley Blackwell, pp.243-254.

Schuwirth, L.W. and Van der Vleuten, C.P. 2011. Programmatic assessment: from assessment of learning to assessment for learning. *Medical teacher.* **33**(6), pp.478-485.

Schwandt, T.A. and Gates, E.F. 2018. Case Study Methodology. In: Denzin, N.K. and Lincoln, Y.S. eds. *The SAGE Handbook of Qualitative Research.* 5th ed. London, UK: SAGE Publications.

Sebai, Z.A. 2019. *The Vision and the Health.* Lebanon: The Arabic Institute for Studies and Publications.

Selby, C., Osman, L., Davis, M. and Lee, M. 1995. How to do it: set up and run an objective structured clinical exam. *Bmj.* **310**(6988), pp.1187-1190.

Setyonugroho, W., Kennedy, K.M. and Kropmans, T.J.B. 2015. Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review. *Patient Education and Counseling.* **98**(12), pp.1482-1491.

Shand, G. 2020. Bringing OSCEs into the 21st century: Why internet access is a requirement for assessment validity. *Medical Teacher.* **42**(4), pp.469-471.

Shanley, C. 2007. Management of change for nurses: lessons from the discipline of organizational studies. *Journal of nursing management.* **15**(5), pp.538-546.

Shannon, B., Shannon, H., Bowles, K.-A., Williams, C., Andrew, N. and Morphet, J. 2022. Health professionals' experience of implementing and delivering a 'Community Care'programme in metropolitan Melbourne: a qualitative reflexive thematic analysis. *BMJ open.* **12**(7), pe062437.

Sharma, C., Singh, P., Shekhar, S., Bhardwaj, A., Jhirwal, M., Ghuman, N.K., Gothwal, M., Yadav, G., Kathuria, P. and Mishra, V. 2023. Pre-COVID and COVID experience of objective structured clinical examination (OSCE) as a learning tool for post-graduate residents in Obstetrics & Gynecology - a quality improvement study. *Obstet Gynecol Sci.* **66**(4), pp.316-326.

Shaw, S., Elston, J. and Abbott, S. 2004. Comparative analysis of health policy implementation: the use of documentary analysis. *Policy Studies.* **25**(4), pp.259-266.

Shea, B.J., Grimshaw, J.M., Wells, G.A., Boers, M., Andersson, N., Hamel, C., Porter, A.C., Tugwell, P., Moher, D. and Bouter, L.M. 2007. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC medical research methodology.* **7**(1), pp.1-7.

Shiffer, C.D., Boulet, J.R., Cover, L.L. and Pinsky, W.W. 2019. Advancing the quality of medical education worldwide: ECFMG's 2023 medical school accreditation requirement. *Journal of Medical Regulation.* **105**(4), pp.8-16.

Shitu, B. and Girma, T. 2008. Objective structured clinical examination (OSCE): examinee's perception at department of pediatrics and child health, Jimma university. *Ethiopian Journal of Health Sciences.* **18**(2), pp.47-52.

Shumway, J.M. and Harden, R.M. 2003. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. *Medical Teacher.* **25**(6), pp.569-584.

Silverman, D. 2017. *Doing Qualitative Research.* 5th ed. London: SAGE.

Simon, S.R., Bui, A., Day, S., Berti, D. and Volkan, K. 2007. The relationship between second-year medical students' OSCE scores and USMLE Step 2 scores. *Journal of evaluation in clinical practice.* **13**(6), pp.901-905.

Simper, N., Mårtensson, K., Berry, A. and Maynard, N. 2022. Assessment cultures in higher education: reducing barriers and enabling change. *Assessment & Evaluation in Higher Education.* **47**(7), pp.1016-1029.

Smirnova, A., Sebok-Syer, S.S., Chahine, S., Kalet, A.L., Tamblyn, R., Lombarts, K.M., van der Vleuten, C.P. and Schumacher, D.J. 2019. Defining and adopting clinical performance measures in graduate medical education: where are we now and where are we going? *Academic Medicine.* **94**(5), pp.671-677.

Smith, L. and Abouammoh, A. 2013. Challenges and Opportunities for Higher Education in Saudi Arabia: An Exploratory Focus Group. In: Smith, L. and Abouammoh, A. eds. *Higher Education in Saudi Arabia: Achievements, Challenges and Opportunities.* Dordrecht: Springer Netherlands, pp.167-179.

Snabe, B. 2007. *The Usage of System Dynamics in Organizational Interventions: A Participative Modeling Approach Supporting Change Management Efforts.* Germany: Deutscher Universitätsverlag.

Solà-Pola, M., Morin-Fraile, V., Fabrellas-Padrés, N., Raurell-Torreda, M., Guanter-Peris, L., Guix-Comellas, E. and Pulpón-Segura, A.M. 2020. The usefulness and acceptance of the OSCE in nursing schools. *Nurse education in practice.* **43**, p102736.

SSME. 2023. *SSME Objectives | Saudi Society for Medical Education.* [Online]. [Accessed 24 June]. Available from: https://ssme.kku.edu.sa/en/content/70

Stake, R.E. 2006. *Multiple Case Study Analysis.* New York: Guilford Publications.

Stalmeijer, R.E., McNaughton, N. and Van Mook, W.N.K.A. 2014. Using focus groups in medical education research: AMEE Guide No. 91. *Medical Teacher.* **36**(11), pp.923-939.

Stassen, M.L.A., Doherty, K., Poe, M., Planning, U.o.M.a.A.O.o.A. and Assessment. 2004. *Program-based Review and Assessment: Tools and Techniques for Program Improvement.* Massachusetts: Office of Academic Planning & Assessment, University of Massachusetts Amherst.

Sterz, J., Linßen, S., Stefanescu, M., Schreckenbach, T., Seifert, L. and Ruesseler, M. 2021. Implementation of written structured feedback into a surgical OSCE. *BMC Medical Education.* **21**(1), pp.1-9.

Stige, B., Malterud, K. and Midtgarden, T. 2009. Toward an agenda for evaluation of qualitative research. *Qualitative health research.* **19**(10), pp.1504-1516.

Styhre, A. 2002. Non-linear change in organizations: organization change management informed by complexity theory. *Leadership & Organization Development Journal.* **23**(6), pp.343-351.

Sudirman, A., Gemilang, A.V. and Kristanto, T.M.A. 2021. The power of reflective journal writing for university students from the EFL perspective. *Studies in English Language and Education.* **8**(3), pp.1061-1079.

Suri, H. 2018. Meta-analysis, systematic reviews and research syntheses. In: Cohen, L., et al. eds. *Research Methods in Education.* 8th ed. Abingdon: Routledge.

Suter, E., Arndt, J., Arthur, N., Parboosingh, J., Taylor, E. and Deutschlander, S. 2009. Role understanding and effective communication as core competencies for collaborative practice. *Journal of interprofessional care.* **23**(1), pp.41-51.

Swanson, D.B., Clauser, B.E. and Case, S.M. 1999. Clinical skills assessment with standardized patients in high-stakes tests: a framework for thinking about score precision, equating, and security. *Advances in Health Sciences Education.* **4**(1), pp.67-106.

Swanson, D.B. and Roberts, T.E. 2016. Trends in national licensing examinations in medicine. *Medical education.* **50**(1), pp.101-114.

Swanson, D.B. and van der Vleuten, C.P. 2013. Assessment of clinical skills with standardized patients: state of the art revisited. *Teaching and learning in medicine.* **25**(1), pp.S17-S25.

Swift, M., Spake, E. and Kohia, M. 2016. Examiner fatigue and ability to concentrate in objective structured clinical examinations for physical therapist students. *Journal of allied health.* **45**(1), pp.62-70.

Talwalkar, J.S., Cyrus, K.D. and Fortin, A.H. 2019. Twelve tips for running an effective session with standardized patients. *Medical Teacher.* **42**(6), pp.622-627.

Tamblyn, R. 1999. Outcomes in medical education: what is the standard and outcome of care delivered by our graduates? *Advances in Health Sciences Education.* **4**(1), pp.9-25.

Tavakol, M. and Dennick, R. 2011a. Making sense of Cronbach's alpha. *International journal of medical education.* **2**, pp.53-55.

Tavakol, M. and Dennick, R. 2011b. Post-examination analysis of objective tests: AMEE Guide no. 54. *Medical Teacher.* **33**(6), pp.447-458.

Tavakol, M. and Sandars, J. 2014. Quantitative and qualitative methods in medical education research: AMEE Guide No 90: Part I. *Medical teacher.* **36**(9), pp.746-756.

Tavakol, M., Sharpe, C.C., Stewart, C. and O'Brien, D.G. 2023. Ranking students in OSCEs: Issues and solutions. *Medical Teacher.* **45**(5), pp.554-554.

Tekian, A. and Almazrooa, A.A. 2011. Does Saudi Arabia need an Abraham Flexner? *Medical Teacher.* **33**(1), pp.72-73.

Tekian, A. and Norcini, J. 2015. Overcome the 60% passing score and improve the quality of assessment. *GMS Zeitschrift für Medizinische Ausbildung.* **32**(4), p43.

Telmesani, A., Zaini, R.G. and Ghazi1, H.O. 2011. Medical education in Saudi Arabia: a review of recent developments and future challenges. *Eastern Mediterranean Health Journal.* **17**(8), pp.703-707.

Terry, G. and Hayfield, N. 2020. Reflexive Thematic Analysis. In: Ward, M.R.M. and Delamont, S. eds. *Handbook of Qualitative Research in Education.* 2nd ed. Cheltenham, UK: Edward Elgar Publishing.

Thornhill, A., Lewis, P. and Saunders, M.N. 1996. The role of employee communication in achieving commitment and quality in higher education. *Quality Assurance in Education.* **4**(1), pp.12-20.

Tijani, K.H., Giwa, S.O., Abiola, A.O., Adesanya, A.A., Nwawolo, C.C. and Hassan, J.O. 2017. A COMPARISON OF THE OBJECTIVE STRUCTURED CLINICAL EXAMINATION AND THE TRADITIONAL ORAL CLINICAL EXAMINATION IN A NIGERIAN UNIVERSITY. *J West Afr Coll Surg.* **7**(3), pp.59-72.

Tracy, S.J. 2010. Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative inquiry.* **16**(10), pp.837-851.

Trigwell, K., Martin, E., Benjamin, J. and Prosser, M. 2000. Scholarship of Teaching: A model. *Higher Education Research & Development.* **19**(2), pp.155-168.

Troncon, L.E.d.A. 2004. Clinical skills assessment: limitations to the introduction of an" OSCE"(Objective Structured Clinical Examination) in a traditional Brazilian medical school. *São Paulo Medical Journal.* **122**(1), pp.12-17.

Tsai, J., Liu, K., Lee, K., Yen, J., Yen, J., Liu, C. and Lai, C. 2008. Evaluation of the effectiveness of postgraduate general medicine training by objective structured clinical examination - pilot study and reflection on the experiences of Kaohsiung Medical University Hospital. *Kaohsiung Journal of Medical Sciences.* **24**(12), pp.627-633.

Turner, J.L. and Dankoski, M.E. 2008. Objective structured clinical exams: a critical review. *J Fam Med.* **40**(8), pp.574-578.

USMLE. 2023. *Work to relaunch USMLE Step 2 CS discontinued | USMLE.* [Online]. [Accessed 16 May]. Available from: https://www.usmle.org/work-relaunch-usmle-step-2-cs-discontinued

van der Vleuten, C. 1996. The assessment of professional competence: Developments, research and practical implications | SpringerLink. *Advances in Health Sciences Education.* **1**(1), pp.41-69.

van der Vleuten, C.P.M., Schuwirth, L., Scheele, F., Driessen, E. and Hodges, B. 2010. The assessment of professional competence: building blocks for theory development. *Best Practice & Research Clinical Obstetrics & Gynaecology.* **24**(6), pp.703-719.

van der Vleuten, C.P.M. and Schuwirth, L.W.T. 2005. Assessing professional competence: from methods to programmes. *Medical Education.* **39**(3), pp.309-317.

van Schaik, S.M., Reeves, S.A. and Headrick, L.A. 2019. Exemplary learning environments for the health professions: A vision. *Academic Medicine.* **94**(7), pp.975-982.

Varpio, L., O'Brien, B., Rees, C.E., Monrouxe, L., Ajjawi, R. and Paradis, E. 2021. The applicability of generalisability and bias to health professions education's research. *Medical Education.* **55**(2), pp.167-173.

Varpio, L., Paradis, E., Uijtdehaage, S. and Young, M. 2020. The distinctions between theory, theoretical framework, and conceptual framework. *Academic Medicine.* **95**(7), pp.989-994.

Vincent, S.C., Arulappan, J., Amirtharaj, A., Matua, G.A. and Al Hashmi, I. 2022. Objective structured clinical examination vs traditional clinical examination to evaluate students' clinical competence: A systematic review of nursing faculty and students' perceptions and experiences. *Nurse Education Today.* **108**, p105170.

Vision2030. 2023. *Saudi Vision 2030.* [Online]. [Accessed 1/3/2023]. Available from: https://vision2030.gov.sa/en

Wahba, H. 1935. *The Arabian Peninsula in the Twentieth Century* Cairo: Dar Alafaaq Alarabia.

Walsh, K. 2014. Documentary research and evaluation in medical education. *Journal of educational evaluation for health professions.* **11**, p18.

Walsh, M., Bailey, P.H. and Koren, I. 2009. Objective structured clinical evaluation of clinical competence: an integrative review. *Journal of advanced nursing.* **65**(8), pp.1584-1595.

Walsh, R. 2003. The methods of reflexivity. *The Humanistic Psychologist.* **31**(4), pp.51-66.

Ware, J., Mardi, A.E., Abdulghani, H. and Siddiqui, I. 2014. *SCFHS OSCE Manual 2014.* Saudi Arabia: Saudi Commission for Health Specialties (SCFHS).

Wass, V. and van der Vleuten, C. 2004. The long case. **38**(11), pp.1176-1180.

Wass, V., Van der Vleuten, C., Shatzer, J. and Jones, R. 2001. Assessment of clinical competence. *The Lancet.* **357**(9260), pp.945-949.

Weiner, W.F. 2009. Establishing a culture of assessment. *Academe.* **95**(4), pp.28-32.

WFME. 2015. *WFME Global Standards for Quality Improvement.* [Online]. The World Federation for Medical Education. [Accessed 31 July 2023]. Available from: https://wfme.org/wp-content/uploads/2015/01/BME_2015_v2.pdf

WFME. 2020. *WFME Global Standards for Quality Improvement.* [Online]. The World Federation for Medical Education. [Accessed 31 July 2023]. Available from: https://wfme.org/wp-content/uploads/2020/12/WFME-BME-Standards-2020.pdf

Whitehouse, A., Hassell, A., Bullock, A., Wood, L. and Wall, D. 2007. 360 degree assessment (multisource feedback) of UK trainee doctors: Field testing of team assessment of behaviours (TAB). *Medical Teacher.* **29**(2-3), pp.171-176.

Wijk, H., Heikkilä, K., Ponzer, S., Kihlström, L. and Nordquist, J. 2021. Successful implementation of change in postgraduate medical education–a qualitative study of programme directors. *BMC Medical Education.* **21**, pp.1-8.

Wilby, K.J., Black, E.K., Austin, Z., Mukhalalati, B., Aboulsoud, S. and Khalifa, S.I. 2016. Objective structured clinical examination for pharmacy students in Qatar: cultural and contextual barriers to assessment. *Eastern Mediterranean Health Journal.* **22**(4), pp.251-257.

Wilby, K.J. and Diab, M. 2016. Key challenges for implementing a Canadian-based objective structured clinical examination (OSCE) in a Middle Eastern context. *Canadian Medical Education Journal.* **7**(3), pp.e4-e9.

Wilkinson, T.J., Newble, D.I., Wilson, P.D., Carter, J.M. and Helms, R.M. 2000. Development of a three-centre simultaneous objective structured clinical examination. *Medical Education.* **34**(10), pp.798-807.

Wood, T.J. and Pugh, D. 2019. Are rating scales really better than checklists for measuring increasing levels of expertise? *Medical Teacher.* **42**(1), pp.46-51.

Yardley, L. 2000. Dilemmas in qualitative health research. *Psychology and health.* **15**(2), pp.215-228.

Yazan, B. 2015. Three approaches to case study methods in education: Yin, Merriam, and Stake. *The qualitative report.* **20**(2), pp.134-152.

Yeates, P., Cardell, J., Byrne, G. and Eva, K.W. 2015a. Relatively speaking: contrast effects influence assessors' scores and narrative feedback. *Medical Education.* **49**(9), pp.909-919.

Yeates, P., Cope, N., Hawarden, A., Bradshaw, H., McCray, G. and Homer, M. 2019. Developing a video-based method to compare and adjust examiner effects in fully nested OSCEs. *Medical Education.* **53**(3), pp.250-263.

Yeates, P., Maluf, A., Kinston, R., Cope, N., McCray, G., Cullen, K., O'Neill, V., Cole, A., Goodfellow, R. and Vallender, R. 2022. Enhancing authenticity, diagnosticity and equivalence (AD-Equiv) in multicentre OSCE exams in health professionals education: protocol for a complex intervention study. *BMJ open.* **12**(12), pe064387.

Yeates, P., Moreau, M. and Eva, K. 2015b. Are examiners' judgments in OSCE-style assessments influenced by contrast effects? *Academic Medicine.* **90**(7), pp.975-980.

Yeates, P., Moult, A., Cope, N., McCray, G., Xilas, E., Lovelock, T., Vaughan, N., Daw, D., Fuller, R. and McKinley, R.K.B. 2021. Measuring the effect of examiner variability in a multiple-circuit objective structured clinical examination (OSCE). *Academic medicine.* **96**(8), pp.1189-1196.

Yin, R.K. 2018. *Case Study Research and Applications: Design and Methods.* 6th ed. London: SAGE Publications.

Yousuf, N., Violato, C. and Zuberi, R.W. 2015. Standard setting methods for pass/fail decisions on high-stakes objective structured clinical examinations: a validity study. *Teaching and learning in medicine.* **27**(3), pp.280-291.

Yudkowsky, R. 2019. Performance tests. In: Yudkowsky, R., et al. eds. *Assessment in Health Professions Education.* 2nd ed. London: Taylor & Francis, p.346.

Yudkowsky, R., Downing, S.M. and Park, Y.S. 2019. *Assessment in Health Professions Education.* 2nd ed. New York: Taylor & Francis.

Zabar, S., Kachur, E.K., Kalet, A. and Hanley, K. 2013. *Objective structured clinical examinations: 10 steps to planning and implementing OSCEs and other standardized patient exercises.* New York: Springer.

Zaini, R.G. 2007. *NATIONAL CONSENSUS OF THE VISION OF THE "SAUDI FUTURE DOCTOR": Current Status and Future Perspective of Medical Education in Saudi Medical Schools* PhD thesis, University of Sheffield.

Zamanzadeh, V., Ghaffari, R., Valizadeh, L., Karimi-Moonaghi, H., Johnston, A.N.B. and Alizadeh, S. 2021. Challenges of objective structured clinical examination in undergraduate nursing curriculum: Experiences of faculties and students. *Nurse Education Today.* **103**, p104960.

Zayyan, M. 2011. Objective structured clinical examination: the assessment of choice. *Oman Medical Journal.* **26**(4), pp.219-222.

Appendices

Appendix 1. Details of the search strategies and the number of articles identified in each database.

The search process carried out in **PubMed**. Results from 1975 to 30 June 2023.

#	Search terms	Results
1	Search ((Objective Structured Clinical Examination[Title/Abstract]) OR OSCE[Title/Abstract]) Sort by: Best Match	3249
2	#1 AND ((((implement*[Title/Abstract]) OR design*[Title/Abstract]) OR quality[Title/Abstract]) OR application[Title/Abstract])	1442
3	#2 AND (medic*[Title/Abstract])	989
4	#3 AND (undergraduate[Title/Abstract])	208
5	#4 limited to English	**197**

The search process carried out in Ovid for the fowling databases: (1) **Embase** 1996 to 30 June 2023, (2) **PsycINFO** 1975 to 30 June 2023, (3) **MEDLINE** 1975 to 30 June 2023, and (4) **Global Health** 1975 to 30 June 2023.

#	Search terms	Results
1	(Objective Structured Clinical Examination[Title/Abstract]) OR (OSCE[Title/Abstract])	2428
2	#1 AND ((((implement*[Title/Abstract]) OR design*[Title/Abstract]) OR quality[Title/Abstract]) OR application[Title/Abstract])	183
3	#2 AND (medic*[Title/Abstract])	35
4	#3 limited to English	**35**

The search process carried out in **Web of Science**. Results from 1975 to 30 June 2023.

#	Search terms	Results
1	(Objective Structured Clinical Examination[Title/Abstract]) OR (OSCE[Title/Abstract])	6511
2	#1 AND ((((implement*[Title/Abstract]) OR design*[Title/Abstract]) OR quality[Title/Abstract]) OR application[Title/Abstract])	2656
3	#2 AND (medic*[Title/Abstract])	1335
4	#3 AND (undergraduate[Title/Abstract])	219
5	#4 limited to English	**207**

The search process carried out in **Cochrane Library**. Results from 1975 to 30 June 2023.

#	Search terms	Results
1	(Objective Structured Clinical Examination[Title/Abstract]) OR (OSCE[Title/Abstract])	1324
2	#1 AND ((((implement*[Title/Abstract]) OR design*[Title/Abstract]) OR quality[Title/Abstract]) OR application[Title/Abstract])	968
3	#2 AND (medic*[Title/Abstract])	896
4	#3 AND (undergraduate[Title/Abstract])	93
5	#4 limited to English	**69**

The search process carried out in **Eric**. Results from 1975 to 30 June 2023.

#	Search terms	Results
1	((title or abstract:"Objective Structured Clinical Examination") or (OSCE))	179
2	#1 AND((title or abstract:"Objective Structured Clinical Examination") or (OSCE)) and (implement* or design* or quality or application)	63
3	#2 AND (TX medic*)	44
4	#3 limited to English	**35**

The search process carried out in **Scopus**. Results from 1975 to 30 June 2023.

#	Search terms	Results
1	(TITLE(objective AND structured AND clinical AND examination) OR T ITLE (OSCE))	2141
2	#1 AND ((TITLE-ABS-KEY implement*) OR TITLE-ABS-KEY (design*) OR TITLE-ABS-KEY (quality) OR TITLE-ABS-KEY application)))	735
3	#2 AND (TITLE-ABS-KEY (medic*))	483
4	#3 AND (TITLE-ABS-KEY (undergraduate))	135
5	#4 limited to English	**131**

PubMed = 197	Ovid (Embase, Global Health, PsycINFO, and MEDLINE) = 35	Web of Science = 207
Cochrane Library = 69	Eric = 35	Scopus = 131
Total = 674		

Appendix 2. The reasons for and numbers of publications excluded at the first level of filtration.

The reasons and number of publications excluded were determined after the title and abstract had been screened.

Number of papers excluded	Reasons for exclusion Not aligned with the primary research question of this scoping review.	Comments
31	One of the following accessibility reasons: • No access to the records. • Records' full text is not available. • Records are not available in English.	
70	The articles that are not about the OSCE. For example, the OSCE is just mentioned as an example somewhere in the abstract.	
52	The articles that are not pertinently related to the OSCE. The OSCE is just mentioned in the abstract for different reasons. For example, discussing clinical assessments in general.	
60	The OSCE is used alone or among other assessment tools to assess a particular subject or course without discussing the OSCE's design or implementation at any level. For example, to examine the effectiveness of a teaching method or make a correlation with other tools.	
36	Articles that are not about medical schools' settings. Papers from other health professions education (HPE), i.e., dentistry, nursing, pharmacy, dietetics, physical therapy, or veterinary medicine. Out of the scope of this study.	
7	Articles about the OSCE in postgraduate education, which are outside the scope of this study.	I found 1 paper useful to include
21	Discuss specific OSCE designs such as Telemedicine OSCE (TeleOSCE), online OSCE, eOSCE, or digital OSCE, which is not the focus of this research.	
277	The total number of excluded articles	

Appendix 3. The reasons, number, and citations of publications were excluded at the second level of filtration.

The reasons, number, and citations of publications excluded after the full-text had screened.

Number of papers excluded	Reasons for exclusion Not aligned with the primary research question of this scoping review.		Citations
23	Not about OSCE, generic papers, or other clinical assessment tools. For example, the objective structured long examination record (OSLER), the standardised orthopaedic assessment tool (SOAT), the interprofessional team objective structured clinical examination (ITOSCE), the team objective structured clinical examination (TOSCE), using OSCE to examine the effect of ethnicity and gender on performance, and other overarching concepts They may mention the OSCE as an example or for other purposes.		(Gleeson, 1997; Bradley and Bligh, 1999; Singleton et al., 1999; Wass et al., 2001; Cullen et al., 2003; Hodges, 2003; Wass et al., 2003; Downing, 2004; Wass and Van Der Vleuten, 2004; Haq et al., 2005; Hamdy et al., 2006; Furman, 2008; Naeem et al., 2012; Lafave and Katz, 2014; Deane et al., 2015; Sureda-Demeulemeester et al., 2017; Świerszcz et al., 2017; Gordon et al., 2019; Lin et al., 2019; Parandavar et al., 2019; Furstenberg et al., 2020; Staziaki et al., 2020; Nourkami-Tutdibi et al., 2021)
38	Publications here have little to no information about the OSCE's design or implementation. They are not discussing the main focus of this study. They may also have a lack of information, which makes interpretation or discussion difficult or impossible.	Papers that only use the OSCE to examine a particular topic, e.g., communication skills, history teaching, the effectiveness of case-based teaching (CBT), cost, comparison between the old and new curriculum, patient safety, cultural competence, and serious games for OSCE training.	(Cater et al., 1991; Lunenfeld et al., 1991; Poenaru et al., 1997; Bradley and Humphris, 1999; Kelly et al., 2002; Beckert et al., 2003; Roberts et al., 2003; Jacobs et al., 2004; Kelly and Murphy, 2004; Alinier et al., 2006; Ishikawa et al., 2006; Buckley and Zamora, 2007; Green et al., 2007; Peeraer et al., 2007; Smith et al., 2009; Amini et al., 2012; Bosse et al., 2012; Kumar and Krishna, 2012; Kiessling and Langewitz, 2013; Mirzazadeh et al., 2013; Piryani et al., 2013; Mohd Saiboon et al., 2014; Chan et al., 2015; Ginsburg et al., 2015; Amini et al., 2016; Bartlett et al., 2016; Brashers et al., 2016; Chahine et al., 2016; Bernard et al., 2017; Butler et al., 2017; Seifert et al., 2017; Yusuf et al., 2017; Heide et al., 2018; Kelly et al., 2018; Gordon et al., 2019; Germa et al., 2021; Hoehne et al., 2022; Nct, 2022)
4		Papers that discuss the OSCE superficially and without giving evidence or justifications for their choices.	(Burge and Lancaster, 2004; Power et al., 2006; Daud-Gallotti et al., 2011; Burt et al., 2016)
5		Papers that only discuss the examiners' effect on scoring. However, these articles were used somewhere else in this thesis.	(Wood et al., 2017; Yeates and Sebok-Syer, 2017; Sturman et al., 2018; Coetzee and Monteiro, 2019; Yeates et al., 2019)
16		Papers that focus on psychometric analysis. For example, calculating the reliability and correlation between OSCE and other assessments. However, these articles were utilised somewhere else in this thesis.	(Famuyiwa et al., 1991; Collins et al., 1995; Martin et al., 2000; Wass and Jolly, 2001; Wilkinson and Frampton, 2004; Power et al., 2006; Rekany and Aldabbagh, 2010; Markey et al., 2011; Fong et al., 2012; Mondal et al., 2012; Niehaus et al., 2012; Pugh et al., 2016; Trejo-Mejía et al., 2016; P. et al., 2017; Tijani et al., 2017; Müller et al., 2019)
8		Papers that discuss the application of sequential OSCE or dealing with borderline student.	(Rashid et al., 2011; Pell et al., 2012; Pell et al., 2013; Young et al., 2014; Mortaz Hejri et al., 2016; Hejri and Jalili, 2017; Homer et al., 2018; Wikander and Bouchoucha, 2018)
17		Papers that only discuss very specific elements of the OSCE that do not inform this scoping review, such as standard-setting, feedback, fairness, or scoring systems (e.g., checklists and global ratings). However, these articles were utilised somewhere else in this thesis.	(McILROY, 2000; Duffield and Spencer, 2002; Cusimano and Rothman, 2003; Chesser et al., 2004; Cusimano and Rothman, 2004; Reiter et al., 2004; Liu and Liu, 2008; McKinley et al., 2008; Managheb et al., 2012; Mortsiefer et al., 2014; Denison et al., 2016; Kamath et al., 2016; Homer et al., 2017; Yusuf et al., 2017; Shulruf et al., 2018; Sterz et al., 2022; Yang et al., 2022; Pinilla et al., 2023)
13		Papers that only discuss students' perception, opinion, peer assessment, or peer-teaching.	(Kowlowitz et al., 1991; McIlroy et al., 2002; Pierre et al., 2004; Tagawa and Imanaka, 2010; Mondal et al., 2012; Khorashad et al., 2014; Labaf et al., 2014; Lafleur et al., 2015; Furmedge et al., 2016; Mamatha et al., 2017; Lee et al., 2018; Muller et al., 2018; Majumder et al., 2019)
12		Papers that only discuss simulated patients, standardised patients, simulation, high-fidelity simulation, online OSCE or video OSCE.	(McGraw and O'Connor, 1999; Srinivasan et al., 2006; Vivekananda-Schmidt et al., 2007; Iramaneerat et al., 2008; Hendrickx et al., 2009; Homer and Pell, 2009; Verma et al., 2011; Leung and Nicholls, 2014; Zeng et al., 2014; Lin et al., 2015; Sabzwari et al., 2017; Kaltman et al., 2018; Guerrero et al., 2023)
136	The total number of excluded articles		

Appendix 4. The reasons, number, and citations of publications were excluded at the third level of filtration.

The reasons, number, and citations of publications excluded after the full-text had been read and assessed for eligibility.

Number of papers excluded	Reasons for exclusion Not aligned with the primary research question of this scoping review.	Citations
15	Articles not strictly referring to and discussing the OSCE implementation, design, or application process or not in medical settings.	(McFaul et al., 1993; Duerson et al., 2000; Wilkinson et al., 2000; Rushforth, 2007; Dadgar et al., 2008; Al-Haddad, 2009; Walsh et al., 2009; Hawker and Walker, 2010; Fuller et al., 2013; Bord et al., 2015; Brown et al., 2015; Shirwaikar, 2015; Wilby et al., 2016; Schleicher et al., 2017; Talwalkar et al., 2019)
14	Articles in the Saudi context but not referring to or discussing the OSCE implementation, design, or application process. They may only be discussing the e-OSCE or online OSCE, which is not the focus of this research. These articles discuss topics similar to the OSCE, so they may mention the OSCE in the article..	(Al Kadi and Salati, 2012; Amr, 2012; Guraya and Almaramhy, 2012; Yaqinuddin et al., 2013; Al-Alwan et al., 2015; Alaki et al., 2016; Hassan et al., 2017; AlShammari et al., 2018; Alshaya et al., 2021; Shaiba et al., 2021; Shorbagi et al., 2022; Yousef et al., 2022; Abdulghani et al., 2023; Alhamad et al., 2023)
29	The total number of excluded articles	

Appendix 5. Publications discussing the OSCE implementation.

#	Citation & Country	Topic - Objective	Design	Key Findings	Relation to my Work - Notes
1	(Harden et al., 1975) UK	The original manuscript of the notion of the OSCE	Research paper	The notion of Objective Structured Examination has been introduced to overcome some disadvantages of the traditional clinical examination	-Set the historical background for the development of the OSCE. -Descript the primary idea of the OSCE.
2	(Harden and Gleeson, 1979) UK	Topic: Assessment of clinical competence using an objective structured clinical examination (OSCE)	Booklet	-The initial guideline developed by the founders of the OSCE. -They discussed the key areas of the OSCE, advantages, disadvantages, and steps of implementing the exam.	-Discerption of the original OSCE -Advantages -Disadvantages -Preparation -This paper shows how the OSCE has been developed over time.
3	(Harden, 1988) UK	Topic: What is an OSCE?	Overview article	-Briefly describing the OSCE -Highlighting the main advantages of the OSCE.	-Overview on the OSCE from the founder -Lack of practical steps of implementation. -Shows the source of error in the OSCE.
4	(Harden, 1990) UK	Topic: Twelve tips for organizing an OSCE.	Recommendations	Harden identified twelve steps and tips for organising an OSCE.	-Some good advice on organising the exam been introduced. -It is a superficial guide.
5	(Harden, 2015) UK	Topic: Misconceptions and the OSCE	Opinion/ Commentary	-Harden provides a solution for the problems facing the OSCE int its feasibility, compartmentalise, authenticity, the role of the examiner, and student stress.	-Harden discussed the most common (five) disadvantages that have been raised in the literature.
6	(Harden, 2016) UK	Topic: Revisiting 'Assessment of clinical competence using an objective structured clinical examination (OSCE)'	Reflection	-Harden mentioned the story behind his innovative way and reflected on the lessons that he learned by introducing the OSCE	-Not all OSCEs are good, therefore, more consideration should be paid for the implementation to achieve high-quality OSCEs.
7	(Harden et al., 2015) UK	Title: The Definitive Guide to the OSCE: The Objective Structured Clinical Examination as a performance assessment	Book	Cover almost all aspects of the OSCE	-Very relevant to my project focus. -A comprehensive guide for the OSCE. -The downside of detailed books that they are not easily approachable by busy clinicians who design and implement the OSCE.
8 & 9	(Boursicot et al., 2011; Boursicot et al., 2020) Ottawa conference	Title: Performance assessment: Consensus statement and recommendations from the 2020 Ottawa Conference	Consensus statement	Provided recommendations for performance assessments from assessment experts	-Historical background -Theoretical description - Recommendations for the OSCE implementation
10	(Abdulghani et al., 2015) Singapore	Title: An Essential Guide to Developing, Implementing, and Evaluating Objective Structured Clinical Examination (OSCE)	Book	Cover almost all aspects of the OSCE	-Very relevant to my project focus. -Some chapters may need further expansion and development e.g., post-doc quality assurance. -The downside of detailed books that they are not easily approachable by busy clinicians who design and implement the OSCE.

11	(Zabar et al., 2013) USA	Title: Objective structured clinical examinations: 10 steps to planning and implementing OSCEs and other standardized patient exercises	Book	-Cover most aspects of the OSCE -The book mostly discusses the organisational aspects of the OSCE.	-It is a relevant book as it mostly focusses on the standardised patients and OSCE. -It is a very concise book, so it did not discuss all key areas of the OSCE implementation
12	(Humphrey-Murto et al., 2013) UK	Objective Structured Clinical Examination in: Oxford Textbook of Medical Education	Book chapter (45)	An overview of the OSCE	-A decent overview of the key component of the OSCE.
13	(GMC, 2023a) UK	Thematic report on the Clinical and Professional Skills Assessment (CPSA) formative meetings	Thematic report	A comprehensive framework to evaluate the OSCE implementation	- Listing the requirements for a high-quality OSCE. - Provide a conscience description for each requirement (design element).
14	(Boursicot and Roberts, 2005)	How to set up an OSCE	Practical steps	-What they did on their OSCE -Recommendations -Suggestions	-Practical steps to implement and ensure high-quality OSCE
15	(Khan et al., 2013b) UAE	The OSCE: AMEE Guide No. 81. Part I: An historical and theoretical perspective	Guide	The theoretical background of the OSCE	-To set the theoretical background for the OSCE. - Show how the OSCE met the criteria of good assessment.
16	(Khan et al., 2013a) UAE;	The OSCE: AMEE Guide No. 81. Part II: Organisation Administration	Guide	Description for designing the OSCE (before, during and after the exam).	-The practical steps of designing and implementing the OSCE. -Sound consensus with other main resources. -Comprehensive and detailed description for the OSCE. -Lack of statistical and psychometric quality measurements.
17	(Pugh and Smee, 2013) Canada	Guidelines for the development of objective structured clinical examination (OSCE) cases	Guideline	They descript 8 steps to develop the OSCE based on the Medical Council of Canada.	-It is useful in blueprinting and writing the cases and designing the scoring sheet. -It does not cover all aspect of the OSCE implementation.
18	(Daniels and Pugh, 2018) Canada	Twelve steps to design an effective OSCE	Guideline	-Useful steps for implementing the OSCE.	-Incomprehensive and very concise guide may help in dome aspect in implementing the OSCE.
19	(Brannick et al., 2011) USA	To assess the OSCE reliability	Systematic review	-The OSCE is not very reliable -Cronbach alpha is 0.66 for 64 studies and 457 reliability values.	-Useful in showing the influence of the OSCE design and construct in calculating its reliability. -Shows the paramount importance of ensuring the quality of the OSCE design and implementation process.
20	(Turner and Dankoski, 2008) USA	OSCE critical review	Review	-Attention should be paid to the implementation aspects to ensure valid and reliable assessment.	-Implementation is critical for the validity of the exam. -The OSCE is not necessarily superior to the other clinical assessment formats.
21	(Barman, 2005) UK	Critique on the OSCE.	Findings of researches published between 1975-2004	-Planning and administration are important to obtain a reasonable level of validity and reliability.	-The implementation is critical to producing high-quality OSCE. -No discussion for the implementation process.

22	(Gupta et al., 2010) India	To discuss validity, reliability, objectivity, and standard-setting of the OSCE	Review and opinion article	-The need for careful design. -The OSCE should not be used alone in the assessment. -It is resource-intensive.	-Factors contributing to a successful OSCE. -Validity, objectivity, reliability, and standard-setting of OSCE
23	(Hodges, 2003) Canada	Validity and the OSCE	Review	-The OSCE is highly contextual and heavily influenced by culture.	-The need for a qualitative and quantitative study to assess OSCE validity in a given place.
24	(Patricio et al., 2009) Portugal	To produce a comprehensive checklist in how a researcher should report studies about the OSCE	Review	-Most papers about the OSCE are lack of info which made them of limited value. -They produce a checklist that contains almost all elements needed to be reported in the OSCE studies.	-Inform me about the quality of the published literature about the OSCE. -Inform my interview and focus group question guide.
25	(Patricio et al., 2013) Portugal	To discuss the feasibility of the OSCE	Literature review	-It is a resource-intensive exam, but it produces a reliable result. -It is recommended exam format to assess clinical competencies.	-The article proof the complexity and expensiveness of the exam, but they suggest it is a worthwhile and best available tool to assess clinical competencies.
26	(Casey et al., 2009) USA	OSCE in Gynaecology	Report	-Overviewing very briefly some implementation aspect for the OSCE.	-The incomprehensive and very concise report may help in some aspects of implementing the OSCE.
27	(Selby et al., 1995) UK	How to set up an OSCE	Overview	A brief overview of how to set up the OSCE	Incomprehensive manual to set the OSCE.
28	(Fuller et al., 2012) UK	Comments on 'A systematic review of the reliability of objective structured clinical examination scores'	Narrative comments	-The need for OSCEs to have adequate construct validity as part of a wider assessment framework. -The important component of quality, careful psychometric analysis at the station level.	-It is likely that much of this variation reflects the diversity of OSCEs themselves in terms of 'usage', design, and construct.
29	(Carraccio and Englander, 2000) USA	Literature review of the OSCE in paediatric	Literature review	-The OSCE should be designed carefully to obtain favourable quality.	-The importance of the OSCE design in producing high-quality exam. -No discussion for the implementation process.
30	(Roberts et al., 2006) UK	To introduce a strategy to improve the quality of clinical assessment	Review/ Reporting experience	-Achieving high-quality OSCE assessment needs long and complex preparation	-Provide valuable experience-based strategy to enhance the OSCE implementation. -Focusing on the importance of examiner training and feedback to the students.
31	(Newble, 2004) UK	Techniques for measuring clinical competence	Review	OSCE is a good alternative for the traditional clinical exams	Longer OSCE is always better than shorter ones.
32	(Gormley, 2011) UK	Constructing OSCE in undergraduate medical education	Guide	-Guidance to develop an OSCE exam. -The OSCE is an important assessment tool -The OSCE should be used in conjunction with other assessment tools. -Quality criteria should be seriously considered.	-Provide a valuable OSCE outline -Discuss some quality criteria of the OSCE -Superficial but useful manual to implement OSCE. -Not comprehensive with no further details
33	(Heal et al., 2019) Australia	To document the similarities and differences between OSCE implementation against gold standard practice	Questionnaire	-There is a wide variation between medical schools in the OSCE practise. -Highlight the need for standardised practice in all aspect of the OSCE.	-Highlight the differences and similarities of the OSCE design and implementation. -Indicate that the OSCE design needs further development and consensus.

34	(Zayyan, 2011) Nigeria	To descript the OSCE	Descriptive article	-It is better than the traditional clinical exam. -Discuss the advantages and disadvantages of the OSCE.	-Concisely review the OSCE. -The OSCE is the best available option to assess students' clinical competencies
35	(Pell et al., 2010) UK	How to measure the quality of the OSCE: A review of metrics	Guide	Discuss the options available to measure the quality of the OSCE.	-Seven metrics have been discussed. -Highlighting the necessity of quality check for the OSCE.
36	(Homer et al., 2017) UK	To discuss the borderline student in the OSCE standard setting	Article	-It is challenging to set the standards for borderline students. -The article highlights the benefit of the sequential OSCE in this regard.	-Sequential OSCE method help in improving the decision regarding the borderline students. -Pre-defined standard-setting methods might be not the optimal methods for OSCE exams.
37	(Pell et al., 2013) UK	Discussing the sequential testing in theory and practise	Article	-Sequential OSCE can increase the reliability for borderline students. -The importance of the quality of the OSCE at the station level.	- The benefits of sequential testing to decrease the cost of the re-set OSCE and enhance OSCE validity and reliability.
38	(Setyonugroho et al., 2015) Indonesia	To assess the ability of the OSCE checklists to assess communication skills	Systematic review	-Lack of agreement between reviewers on the competencies that need to be observed during communication skills OSCEs.	-They recommend produce and use standardising scoring checklist which would enhance the validity of the scores.
39	(Yeates et al., 2019) UK	Developing a video-based method to compare and adjust examiner effects in fully nested OSCEs	Empirical study	-The examiner cohorts influence students' scores in the OSCE.	-The effect of the examiner on students score is inevitable. -Emphasising the need for adequate examiner training and sampling.
40	(MacLean et al., 2018) Australia	To evaluate SP training on exam reliability	Experimental study	-Their findings reveal the superior reliability among SPs who received training.	-The importance of training the SPs to get high inter-rater reliability.
41	(Vleuten et al., 2010) Netherlands	The assessment of professional competence	Theoretical article	-Evaluate the assessment format is important to assess its effect on the learners. -No one assessment can assess all facets of any domain.	-Multiple assessment types should be used to get built-in validity. -Examiner bias is inevitable. -Qualitative judgment on students' performance could be enough.
42	(Vleuten and Schuwirth, 2005) Netherlands	Assessing professional competence: from methods to programmes	Theoretical article	-Professional judgment based on qualitative information is required. -The results of the OSCE should be combined with other assessment to make a good judgment on students' performance.	In designing the OSCE more attention should be paid for its validity, reliability, educational impact, and acceptability.
43	(Bearman et al., 2021) Australia	The hidden labours of designing the Objective Structured Clinical Examination: a Practice Theory study	Theoretical description	To detail how the OSCE can be feasible and sustainable by detailing its implementation processes	A very comprehensive and practical guide to implementing a feasible OSCE.
44	(Heal et al., 2019) Australia	A snapshot of current OSCE practice at Australian medical schools	Descriptive account – questionnaire	To compare the current OSCE practices at 19 Australian medical schools	It shows the value of collaboration between medical schools in implementing the OSCE.
45	(Boursicot et al., 2022) International	Quality assurance of OSCE	Book chapter	Provide an illustration for the application of Kane's validity framework	Show how the validity of the OSCE can be evaluated using a contemporary validity theory.

Appendix 6. Papers discussing the OSCEs implemented in new locations.

#	Citation Country	Topic - Objective	Design	Key Findings	Relation to my Work - Notes
1 & 2	(Abdelaziz et al., 2016; Sallam et al., 2022) Egypt	To evaluate medical school in Egypt in designing and implementing the OSCE for the first time	Quantitative (Questionnaire for students and assessors & psychometric analysis)	-The OSCE is recourse intensive. -They included somewhat good details for the implementation process. -The reliability was acceptable.	-The evaluation was merely quantitative. -Other stake holder's opinion not included. -No suggestions for improvements.
3	(Iqbal et al., 2009) Pakistan	Implementing the OSCE in resource-limited context	Quantitative/survey	-Successfully implemented. -Raise awareness for the need for more clinical assessments. -Their experience encouraged other departments to adopt the OSCE.	-It shows how they overcome restricted resources. -Overall satisfaction among students and staff.
4	(Saeed et al., 2012) Pakistan	To assess the validity of their newly implemented OSCE	Quantitative/survey & psychometric analysis	-The need for multi-source evidence to assess the quality of the OSCE. -They can point out some source of error in their assessment.	-They highlight the need for multiple pieces of evidence to support the assumptions about the quality of the OSCE. -Not enough details about the implementation process.
5	(Troncon, 2004) Brazil	To assess their experience with the OSCE from students and faculty point of view	Descriptive - semi-quantitative study	-Students report it is stressful and they critic its arrangements. -Faculty report it is compartmentalised approach to the patients, critic its feasibility, and its educational impact. -They faced many hindering forces	-It shows the limitations that traditional school suffer from with the OSCE. -The negative impact of the absence of supportive educational climate which may seem relevant to our context. -The experience of a medical school that newly adopt the OSCE.
6	(Bhatnagar et al., 2011) India	To disseminate the usefulness of OSCE in clinical assessment	Reporting experience	-The validity and reliability of OSCE in assessing students' clinical skills. -It is a resource-intensive method.	-It shows the need for the OSCE. -Lake of info about the design and implementation process. -The need for motivation to maintain the OSCE in our assessment toolkit.
7	(Khapre et al., 2020) India	To investigate faculty opinion on the OSCE	Qualitative survey	-Barriers due to the lack of support from the faculty. -Low utility due to low validity and reliability.	The need for faculty development sessions and the administrative support.
8	(Huang et al., 2007) Taiwan	To descript their experience in design and implement the OSCE in their medical school	Descriptive and commentary + Questionnaire	-Useful assessment -Some challenges -Export the expertise to other departments	Show how they improve the OSCE from scratch to success.
9	(Lee et al., 2008) Taiwan	To follow up on the previous paper	Descriptive	-Overview the journey of their PSCE from 1992 till now. -Illustrate how they improve their OSCE.	-The impact of the supportive education environment -The process that they follow to develop their OSCE.

10	(Tsai et al., 2008) Taiwan	To evaluate the effectiveness of their OSCE for postgraduate year-one residents	Quantitative + Qualitative	-The OSCE is effective and generally perceived well from students and tutor but not examiners. -It is feasible and effective mean of assessment that needs further development and improvement.	-This paper interestingly shows the development and follow up application of OSCE from the same college of medicine and its teaching hospital. -It highlights the importance of the research and follow-up in improving the assessment tools.
11	(Onwudiegwu, 2018) Nigeria	To exploring the OSCE and their recent experience with implementation and provide an easily understood to the OSCE document	Literature search/Descriptive	Producing a document describing the OSCE and how to implement it in a new environment	-Kind of a general review to the OSCE -Exploring the experience of implementing the OSCE in a new place -The paper was directed to the West African context -It includes very little about their own implementation
12	(Sainterant et al., 2019) Haiti	To report how the OSCE been introduced to Haiti medical education	Short descriptive article	-They report an overall positive experience despite some challenges. -The highlight the need for improvement.	-Indicate how the OSCE is flexible as being able to be introduced in a very low-income setting.
13	(Karam et al., 2018) Lebanon	To evaluate the validity of their newly implemented OSCE in a newly established medical school	Psychometric analysis	-They suggest the need for continuous and iterative psychometric analysis to ensure exam validity. -They got a low-reliability OSCE.	-Reporting their experience with the OSCE. -Rather than the psychometric analysis, they add very little details to tell about their OSCE implementation quality.
14	(Shitu and Girma, 2008) Ethiopia	To assess the examinee perception of the OSCE	Questionnaire	-An overall positive attitude of students to this exam format. -Standardised implementation procedure should be strictly followed. -continuous improvement is required.	-The acceptability of students to this new assessment format. -The need for timely feedback.
15	(Ataro et al., 2020) Ethiopia	To explore the experience and challenges of an Ethiopian medical school with the OSCE	Qualitative (interviews)	-Poor organisation and the low number and duration of the stations. -OSCE is acceptable to the examiners and students.	-The OSCE is challenging to design in new places. -It required good resources and preparation
16	(Wilby and Diab, 2016) Qatar	To highlight the challenges of implementing the OSCE in Qatar	Descriptive account	There are some cultural and contextual challenges between the OSCE used in Western and Middle Eastern settings.	-The need for trained SPs. The assessor needs calibration and training. -The standard setting needs to be changed to BRM.
17	(Alkhateeb et al., 2022) Iraq	To evaluate OSCE challenges and opportunities – students' perspective	Questionnaire	-The students believe the OSCE is not fair. -The examiners' behaviour needs to be improved. -The duration of the stations needs to be longer.	-The OSCE organisation is challenging. -The need for examiners and SPs training.

Appendix 7. Publications discussing the OSCEs in SA.

All articles from Saudi literature whose titles or abstracts contain the term OSCE. I have read the full texts of all 44 articles and coded them according to their relevance to my research topic (see colour coding below). All papers published until 30 June 2023, are included.

Useful = 4 (Papers 1-4)	Very little benefit = 19 (Papers 5-30)	Not useful = 14 Excluded (Papers 31-44)

#	Citation	Topic - Objective	Design	Key Findings	Relation to my Work - Notes
1	(Hijazi and Downing, 2008) Medicine/ Postgraduate	To provide an overview and practical guide to the use of the OSCE in the resident's assessment	Concise guide	-OSCE is better than long case assessment.	Even though this paper is from a postgraduate medical education, it adds some practical considerations for implementing high-quality OSCE.
2	(Ware et al., 2014) Generic/ Postgraduate	To describe how to plan, construct and deliver an OSCE in postgraduate medical training programmes	Concise manual	This manual has been written by the Saudi Commission for Health Specialties OSCE team and directed for postgraduate education and testing. It is a short guideline that developed locally to construct an OSCE according to the Saudi Commission standards.	-It describes some practical steps in implementing the OSCE. -It is more in-line with postgraduate education. -Not comprehensive as some important aspects are overlooked such as feedback for the students.
3	(Alnaami, 2008) Medicine/ Undergraduate	To report their experience with their OSCEs and evaluate its validity, reliability, and feasibility	Psychometric analysis/ Stakeholders' opinions	-They report high validity, moderate reliability, and low feasibility -Feasibility needs further work on it.	-Discussing the validity, reliability, and feasibility of their own OSCE of their surgical clerkship. -Well-established schools can afford high demanding OSCE. -The benefit of staff training.
4	(Alnaami et al., 2011) Medicine/ Undergraduate	To improve the psychometric properties of their OSCE and elevate staff motivation	Psychometric analysis/ Stakeholders' opinions	-Their OSCE improved in all source of evidence for its quality. -They reported the positive role of the good arrangements and the examiners being more enthusiastic.	-Discussing the improvement in reliability and validity of their own OSCE of their surgical clerkship. -They add some little description to how they are implementing OSCE. -The importance of staff internal motivation.
5	(Abdullatif, 1992) Medicine/ Undergraduate	To measure the correlation of the OSCE Results with traditional clinical examination and to obtain students perception	Quantitative (Psychometric analysis + questionnaire)	-Their OSCE and long case assessment have a strong positive correlation. -Students opinions reveal high acceptability for the OSCE.	-First reported paper about the OSCE in SA. -The OSCE was utilised among five other assessment formats. -Very superficial without enough details.
6	(Inayah et al., 2017) Medicine/ Undergraduate	To explore the correlation between different types of assessments	Psychometric analysis	-A strong correlation between mid-term OSCE and the final OSCE.	-A little number of OSCE station to make a valid conclusion. -No discussion or discerption for how OSCE is implemented or designed.
7	(Bakhsh et al., 2009) Medicine/ Undergraduate	To compare whither OSCE is as good as the traditional oral clinical examination	Psychometric analysis	-The strong correlation between their OSCE and their traditional clinical examination in assessing surgical course students was found. -The OSCE can be used instead of the traditional clinical examination.	-A very little description of OSCE implementation is provided. -The OSCE is a bit better in its validity than the traditional oral clinical examination.
8	(Moeen-uz-zafar et al., 2015) Medicine/ Undergraduate	To evaluate the reliability and validity of their OSCE	Quantitative (Psychometric analysis + questionnaire)	-Students, internal examiners, and external examiners have been captured. -It correlates weakly with their MCQ. -They claim it valid and credible with weak evidence!	-A very little description for OSCE implementation process is provided. -Their OSCE seems to be stressful but received well by students.

9	(Amr and Amin, 2012) Medicine/ Undergraduate	To compare between assessment formats used in their Psychiatry course	Questionnaire	-Their OSCE correlate will with their traditional oral examination.	-The OSCE has been discussed in very little detail.
10	(Bakhsh et al., 2005) Medicine/ Undergraduate	To detail the logistics of preparation the OSCE	Report	-They fairly detailed the OSCE preparation logistic. -They discovered the low feasibility of it.	-Some fair steps to be notices during preparation for the OSCE. -Highlight the importance of the OSCE bank. -The useless of making inactive station so they are deciding to conduct different exam format for them.
11	(Elfaki et al., 2008) Medicine/ Undergraduate	To share and exchange seven experiences at all levels of applying OSCE	Report	-It is a resource-intensive assessment format. -Due to its limitations, it should be combined with other assessment tools. -They provide basic guidance for newcomers.	-The need for internal motivation and enthusiasm to maintain its implementation. -The OSCE compartmentalise the patient.
12	(Alsaid and Al-Sheikh, 2017) Medicine/ Undergraduate	To explore student and faculty perception on the OSCE of their medical school	Questionnaire	-Student and faculty consider OSCE as a fair assessment tool. -Staff think OSCE is better than traditional long case clinical examination. -Overall positive experience with their OSCE.	-The novelty of the OSCE in a well-established medical school! -Implementation quality can determine the fairness of this exam. -The educational value of the OSCE in highlighting the defect in teaching.
13	(Alghamdi et al., 2016) Medicine/ Undergraduate	To evaluate students' perception and acceptance for the OSCE	Questionnaire	-Students have concerns about inter-rater and inter-patients' variabilities. -Some students believe its stressful and unfair assessment.	-Time allocated for stations should be tailored carefully. -Students need assurance about the fairness and standardisation of the exam.
14	(Elfaki and Al-Humayed, 2016) Medicine/ Undergraduate	To explore the students' acceptance of OSCE	Cross-sectional study	-Students find the OSCE fair and acceptable assessment method.	-Superficial and not useful. -A very little description of OSCE implementation is provided.
15	(Ibrahim et al., 2015) Medicine/ Under- & Post-graduate	To determine student perception towards their OSCE	Cross-sectional study	-The OSCE is an acceptable assessment tool among their students.	-Students have a positive attitude toward the OSCE. -A very little description of OSCE implementation is provided.
16	(Ismail, 2001) Medicine/ Undergraduate	To obtain students perception of the OSCE of fifth-year medical students	Questionnaire	-Positive student opinions. -Good educational impact. -The time and number of stations need to be increased.	-Based on two stations only! -Not clear how these two stations being implemented to evaluate their work.
17	(Raheel and Naeem, 2013) Medicine/ Undergraduate	To assess students' perception of their OSCE	Questionnaire	-Overall, students consider the OSCE as an accepted and fair assessment tool. -Students' feedback is regarded as a key indicator for a successful implementation!	-No discussion or description of how OSCE is implemented or designed.
18	(Al-Eidan et al., 2016) Medicine/ Undergraduate	To examine their course assessment's reliability and validity	Questionnaire	-Students were generally pleased with the OSCE in that course.	-No discussion or description of how OSCE is implemented or designed.
19	(Alaidarous et al., 2016) Medicine/ Postgraduate	To evaluate internal medicine residents' perception of formative OSCE	Questionnaire + Open-ended questions	-Residents have positive attitudes toward the OSCE. -Residents highly appreciate examiners feedback.	-They add some information about the design of that OSCE, but it more in-line with formative and postgraduate settings.

20	(Khairy, 2004) Medicine/ Undergraduate	To assess the acceptability and feasibility of the OSCE for a large number of students	Questionnaire	-Positive attitude toward the OSCE from both students and examiners.	-Highlight the need to expand its use. -It is resource-intensive.
21	(Al-Rubaish et al., 2005) Medicine/ Undergraduate	To collect evidence from assessment tools to improve curriculum delivery	Questionnaire	-The OSCE is under-used, so all departments should consider using it. -They encourage faculty to establish a medical education unit.	-They promote using OSCE in all department. -They suggest providing feedback for students more frequently.
22	(Bajammal et al., 2008) Medicine/ Postgraduate	They are arguing for the need for a Saudi national medical licensing examination that consists of two parts: part I (Written) and part II (OSCE)	Opinion report/ discussion	The authors discuss the need for a national medical licensing examination in Saudi Arabia. They offer OSCE as a possible assessment tool suitable for the proposed exam (part two of that exam).	-Raised the need for a national exam that including OSCE part. -OSCE has not been discussed nor its implementation or design.
23	(Hadi et al., 2018) Pharmacy/ Undergraduate	To assess the association between test-related anxiety and pharmacy students' performance in the	Questionnaire	Students' exam-anxiety may lead to underperformance.	-Only 25 pharmacy students. -No discussion or discerption for how OSCE is implemented or designed.
24	(Alnahdi et al., 2021) Medicine/ Undergraduate	To identify the effect of taking history in Arabic from Arab patients.	Questionnaire	The students prefer to conduct the OSCE in Arabic for history teaching stations..	-It shows the impact of the English language on communication during the OSCE stations. -The student is in favour of the use of the Arabic language..
25	(Alsulimani et al., 2020) Medicine/ Postgraduate	To evaluate the trainees and faculty about using multimodality in the OSCE.	Questionnaire	The participants are in favour of using multiple mobilities in OSCE.	The stakeholders are welcoming innovation and new types of OSCE.
26	(Ansari et al., 2021) Medicine/ Undergraduate	To assess the acceptability of the OSCE among medical students	Questionnaire	The need for better OSCE implementation to improve its acceptability.	The need for better OSCE implementation to improve its acceptability.
27	(Arain, 2021) Medicine/ Undergraduate	To determine the prevalence of anxiety among medical students.	Questionnaire	The OSCE's mobility is increasing anxiety in students.	The acceptability of the OSCE might be attenuated due to its high anxiety rate.
28	(Badawi et al., 2023) Medicine/ Undergraduate	To explore the effectiveness of students scoring the OSCE as an educational tool.	Cross-sectional study	The students benefit from a scoring video showing the blood transfusion process.	-The value of the OSCE as an educational tool. -There is not enough information about the implementation process.
29	(Rajeh et al., 2020) Medicine/ Undergraduate	To conduct contextual analysis to assess the feasibility of introducing a leadership course.	Cross-sectional study	The implementation of a new course might face several contextual challenges..	-It shows the contextual challenges, such as human resources.. -It shows the importance of the OSCE in new courses.
30	(Abu-Zaid et al., 2020) Medicine/ Postgraduate	To argue for the need for a national clinical exam after medical schools.	Theoretical discussion	The authors suggest adding a national clinical part to the current AKT exam.	It highlights the need for quality OSCE to ensure that the students are safe to practise.
31	(Guraya and Almaramhy, 2012) Medicine/ Undergraduate	To evaluate the effectiveness of small group teaching and acquisition of knowledge and skills (Effectiveness of teaching)	Statistical analysis	The OSCE was used among other assessments to evaluate the effectiveness of small group teaching in one course.	This paper has no data about the design or implementation of the OSCE.

32	(Yaqinuddin et al., 2013) Medicine/ Undergraduate	To describe the Objective Structured Practical Examination OSPE and differentiate it from the OSCE. (OSPE)	Report	The article discusses the appropriateness of using OSPE as an assessment tool for Anatomy. The author highlights the differences between OSPE and OSCE. This paper is out of the scope of my research.	This paper is about the OSPE and has no data about the design or implementation of the OSCE.
33	(Al Kadi and Salati, 2012) Medicine/ Undergraduate	To evaluate students' awareness for hand hygiene (Hand hygiene)	Questionnaire	It is not about OSCE. OSCE is only used to examine a particular subject (e.g., hand hygiene practise among students) without describing how OSCE is design, implemented, or evaluated.	This paper has no data about the design or implementation of the OSCE.
34	(Alaki et al., 2016) Dentistry/ Undergraduate	Multiple Mini-Interview MMI reliability in selection for dentistry school (MMI)	Statistical analysis	This paper is in HPE (dentistry). It is not really about OSCE, but it might be related e.g., how they are evaluating the multiple mini-interviews (MMI). The MMI has some similarities to the OSCE.	-Not about OSCE, they just mention it in the abstract as some people consider MMI as a type of OSCE.
35	(Al-Alwan et al., 2015) Medicine/ Undergraduate	To evaluate the relationship between students' perception about the course organiser and their scores in the assessments of that course (Course coordinator behaviour)	Statistical analysis	Not relate, the OSCE process is not discussed at all.	This paper has no data about the design or implementation of the OSCE.
36	(AlShammari et al., 2018) Medicine/ Postgraduate	To compare students' confidence in pre- and post-course, by using OSCE (Simulation course)	Questionnaire	This study is designed to explore the effects of simulation training on paediatric residents. OSCE was used as one method to evaluate that course. No focus has been given for how OSCE is implemented or designed in their course.	This paper has no data about the design or implementation of the OSCE.
37	(Amr, 2012) Medicine/ Undergraduate	To assess the difference between male and female students across different assessment formats (Gender differences in assessment)	Questionnaire	The researcher examines the gender difference in the performance of different assessment tools.	This paper has no data about the design or implementation of the OSCE.
38	(Hassan et al., 2017) Medicine/ Undergraduate	To assess the effectiveness of outpatient clinical teaching of obstetrics and gynaecology (Effectiveness of teaching)	Statistical analysis	This study aimed to test the outpatient clinical teaching. The outcome was measured by OSCE. No discussion or discerption for how OSCE is implemented or designed.	This paper has no data about the design or implementation of the OSCE.
39	(Abdulghani et al., 2023) Medicine/ Undergraduate	To evaluate the value of the Saudi-MEDs framework on students learning (Saudi-MEDs)	Statistical analysis	The Saudi-MED framework made students acquire and feel confident about their clinical competencies.	This paper has no data about the design or implementation of the OSCE.
40	(Alhamad et al., 2023) Pharmacy/ Undergraduate	To discuss the effect of the COVID-19 pandemic on pharmacy education (COVID-19 pandemic)	Book chapter	They use the online OSCE to assess students' performance during the pandemic.	It is irrelevant as it discusses the online OSCE.
41-44	(Alshaya et al., 2021; Shaiba et al., 2021; Shorbagi et al., 2022; Yousef et al., 2022) Medicine/ Undergraduate	To describe the feasibility of conducting the OSCE during the pandemic (COVID-19 pandemic)	Descriptive reports	It discusses the OSCE implementation in a very specific and unusual scenario.	The data included is not relevant to the focus of the research.

Appendix 8. The main elements discussed in the literature that have to be considered in each OSCE to implement a high-quality examination. These elements may have alternate names in the literature, as shown between brackets.

	Main elements	Sub-elements/tasks	Further elements/tasks to be considered
OSCE workforce	**OSCE planning committee (Exam committee/exam panel/organising committee/exam board)**	• Assessment expert/s (e.g., senior assessor) • Educational expert/s (e.g., medical educationist) • Content expert/s (e.g., consultant) • OSCE designer/s	✓ Planning and taking decisions ✓ Deciding timeframe (timeline) ✓ Arranging exam committee meetings
	OSCE designer (organiser/coordinator/ developer/planner/lead)	• Lead the OSCE implementation process • Follow up with all parties • Ensure standardisation • May need co-organiser	✓ Ensure all resources are allocated ✓ Map stations circuit (directions) ✓ Provide briefing and debriefing for all stakeholders
	Stations (scenarios) writers (developer)	• Writing the stations (scenarios) • Writing the instruction (for examiners, students, and patients/SPs) • Inventory for stations requirements	✓ Recruit them early enough ✓ Train them professionally ✓ Peer review for the stations ✓ Piloting the new stations before the actual implementation
	Examiners (assessors/raters)	• Internal examiners • Outside examiners • Consider their experience with the OSCE • Consider their familiarity with the station's content (their speciality) • Send as much exam information as possible before the exam.	✓ Recruit them early enough ✓ Train them professionally ✓ Briefing and debriefing ✓ Calibration or agreement on marking criteria ✓ Allocate reserve examiners
	Patients	• Simulated patients (SPs) • Standardised patients • Real patients	✓ Recruit them early enough ✓ Train them professionally ✓ Briefing and debriefing ✓ Allocate reserve SPs ✓ Pay them
	Evaluators	• External examiners • Psychometricians	✓ Invitations ✓ Briefing and debriefing
	Supporting staff	• Timekeepers (bellman) • Staff to maintain exam security • Staff to oversee the quarantine place	✓ Briefing and debriefing
	Examinees (candidates/students)	• Level • Number • Train and familiarise them with the OSCE process	✓ List of candidates ✓ Briefing
OSCE design	**Intended use of the OSCE**	• OSCE purpose (summative or formative) • The weight of this assessment on the assessment programme	✓ High- or low-stake ✓ To make decisions about students' progress
	Blueprinting (sampling/content mapping/grid)	• Sampling curriculum content • Decide station type • Decide the number of stations • Decide the duration per station • Total exam time	✓ Prepare a list of proposed stations ✓ Peer review
	Marking (scoring) scheme (schedule/ template/rubric/ guidance/sheet/ instrument)	• Checklists • Rating /Likert scale (with scoring rubric) • Weighting	✓ With or without global (overall/holistic) rating (mark/grade)

	Feedback to students	• Develop a built-in system to provide individual written feedback to each student at each station.	✓ Give students the opportunity to receive personal clarification on the feedback they have received.

	Standard setting	• Choose between:
		  ○ Criterion-referenced
		  ○ Norm-referenced
			✓ Criterion-referenced could be:
			-Angoff/Modified Angoff
			-Borderline Regression Method (BRM) -Ebel -Bespoke
			✓ Norm-referenced could be:
			-Cohen -Wijnen

	Recourses	• Fund • Venue • Equipment (e.g., beds, furniture, manikins, and tools) • Preparation time (1-6 months) ahead	✓ Allocate enough fund ✓ Prepare an appropriate venue: -Hospital ward -Clinical skills unit or simulation centre -Large hall
	Logistics	• A safe place for students' smart devices • Measures for quarantine system and place • Catering and refreshments	✓ Directions arrows ✓ Card for station number
	Parallel circuits	• Control set-up • Ensure standardisation	✓ Ensuring a certain level of standardisation across different locations and cohorts (e.g., morning and noon circuits and different sites)
Quality assurance and follow up	**Feedback**	• Receive and give feedback from different stakeholders	✓ Designers ✓ Examiners ✓ Examinees ✓ SPs and real patients
	Evaluation	• Psychometric analysis • External reviewer/evaluator • Internal reviewer/evaluator • Ensuring the educational impact • Informing the curriculum and teaching	✓ The quality metrics could be: -Cronbach's alpha -Number of failures -Between-group variation -Coefficient of determination (R^2) ✓ Calculating the correlation with other exams ✓ Benchmarking with other schools
	Results	• Results announcement system • Implement a complaints and appeals system	✓ Establish a plan to reset failed students
	OSCE bank	• Add valid and reliable stations	✓ Collaboration with other schools

Appendix 9. The documented beginning of OSCE use in Saudi Arabian medical schools.

#	Medical College	Year
1	King Khalid University College of Medicine	1987 (When it was the branch of the King Saud University in Abha)
2	Umm-Alqura University College of Medicine	2000
3	King Abdulaziz University College of Medicine	2003
4	King Saud University College of Medicine	2005
5	Imam Abdulrahman Bin Faisal University College of Medicine	2013

Appendix 10. Focus Group Guide: Designers and writers focus group.

I also designed a slightly different question guide for the examiners' FGs and a third one for the semi-structured interviews.

FG #: _____ How many participants: _____

Medical school: A or B Their job titles: _____

Date: _____ Time: _____ Duration: _____

---

*Thank the participants for volunteering to participate in this study.

*Introduce myself to the participants, what this study is about, and what I plan to do with it.

*Ask the participants for the signed consent forms and explain the focus group ground rules.

A- Opening/introductory questions:

1. What types of clinical assessments do you use to assess your students?
   (summative and formative) (to understand the school assessment programmes).
2. Please tell me how OSCE is conducted at your college.
3. What are your roles in the implementation of OSCE at your college?
4. How would you evaluate OSCE as a clinical performance assessment method?
   (i.e., advantages and disadvantages, and is it an appropriate tool to assess clinical competencies?) (exploring their perspective).

B- Exploring key content questions:

5. What stages do you go through to conduct your OSCEs (A–Z)?
   Please share more information about the following:
   a. Blueprinting stage;
   b. Question and instruction writing;
   c. Designing the scoring rubrics or checklists;
   d. Standard setting;
   e. Recruiting the examiners and SPs;
   f. Arranging the exam venue and tools required;
   g. Exam security and logistic support on the day of the exam.
6. What training did you receive and provide to OSCE examiners and SPs?
7. What affects your choices during the different implementation stages?
   (e.g., evidence-based practice guidelines or frameworks, personal experience, or comfortable choices).
8. What are the challenges in OSCE implementation at your college?
9. What are the opportunities for OSCE implementation at your college?
   For Qs 8&9: Please provide examples.
   (e.g., college regulations, designers, writers, examiners, organisers, SPs, facilities and expenses).
10. Do you think the examiners and SPs perform as they should during the exam?
    (e.g., standardisation of examiners and SPs).
11. What do you think about the quality and appropriateness of your OSCEs overall? Why?
    a. Do you employ any quality assurance measures? What are they?
       (e.g., post-hoc psychometric analysis).
    b. What are the strengths and weaknesses of your OSCE implementation?
    c. Do you think your OSCE has an educational impact besides its assessment value? What impact?
       (e.g., providing students feedback and improving students' clinical skills).
    d. Do you think your OSCEs affect the students' real performance in hospitals? Why?
12. How do the college curriculum and learning activities affect students' performance on your OSCEs?
13. What do you consider to be the role of the college and teachers in preparing students for OSCEs?
14. How would you conduct OSCEs if you had more resources and authority?

C- Closing questions:

15. Do you want to elaborate on or clarify anything?
16. Do you have anything that you want to add or ask?

Thank you for your time. I appreciate your participation.

Appendix 11. The ethical approval from the University of Leeds.

From: Rachel De Souza [Medicine]
Sent: 04 March 2020 12:48 PM
To: Khaled Almisnid [RPG]
Cc: Medicine and Health Univ Ethics Review; Matthew Homer
Subject: Re: MREC 18-103 Study Approval with Comment
Importance: High

Dear Khaled

MREC 18-103 - Exploring Implementation of the Objective Structured Clinical Examination (OSCE) in Saudi Medical Schools

With apologies from the committee for the delay, I am pleased to inform you that the above research ethics application has been reviewed by the School of Medicine Research Ethics Committee (SoMREC) Committee and on behalf of the Chairs, I can confirm a favourable ethical opinion based on the documentation received at date of this email.

The committee suggested you may wish to action the following in the Participant Information Sheet, although this is not a condition of approval:

- **The name of and contact details of supervisors could go near the statement on who has approved this application**
- **Stating who to contact for the interviews should be only to you as the applicant, and not all your supervisors as well**

Please retain this email as evidence of approval in your study file.

Please notify the committee if you intend to make any amendments to the original research as submitted and approved to date. This includes recruitment methodology; all changes must receive ethical approval prior to implementation. Please see https://leeds365.sharepoint.com/sites/ResearchandInnovationService/SitePages/Amendments.aspx or contact the Research Ethics Administrator for further information (FMHUniEthics@leeds.ac.uk) if required.

Ethics approval does not infer you have the right of access to any member of staff or student or documents and the premises of the University of Leeds. Nor does it imply any right of access to the premises of any other organisation, including clinical areas. The committee takes no responsibility for you gaining access to staff, students and/or premises prior to, during or following your research activities.

Please note: You are expected to keep a record of all your approved documentation, as well as documents such as sample consent forms, risk assessments and other documents relating to the study. This should be kept in your study file, which should be readily available for audit purposes. You will be given a two week notice period if your project is to be audited.

It is our policy to remind everyone that it is your responsibility to comply with Health and Safety, Data Protection and any other legal and/or professional guidelines there may be.

I hope the study goes well.

Best wishes
Rachel
On behalf of Dr Naomi Quinton and Dr Anthony Howard, co-Chairs, SoMREC

~~~~~~

Rachel de Souza, Lead Research Ethics & Governance Administrator, The Secretariat, Room 9.29, Level 9, Worsley Building, Clarendon Way, University of Leeds, LS2 9NL, Tel: 0113 3431642, r.e.desouza@leeds.ac.uk
~~~~~~

Appendix 12. A copy of the emails that I sent to the two medical school deans.

Dean Invitation Letter

Dear Dean of [name of the medical school],

I hope you and your colleagues are well.

I am conducting a research project in my PhD study of medical education at the University of Leeds. The research aims to explore the implementation of the OSCE in the Saudi context and develop recommendations for improvement, which medical schools can use to improve their OSCE implementation and produce more defensible results.

I have selected your medical school because I believe that your experience is valuable to the research. The study requires a document analysis of school assessment documents and conducting focus groups and interviews with school staff to provide a better understanding of the context and system of assessment at your school. If the school permits me to conduct the study, I would like to have the documents mentioned in the attached information sheet. Also, I would like to have the names and contact emails of the following staff members:

- All staff involved in the OSCE (i.e. OSCE designers, writers, and examiners).
- Academic leaders (i.e. assessment unit head, medical education unit head, and vice dean for academic affairs).

Further information regarding the study can be found in the information sheet. This study has received ethical approval from the University of Leeds. I have included my contact details below. Please feel free to contact me anytime with any questions or concerns you might have.

I greatly appreciate your support and cooperation.

Kind Regards,

Khaled Mohammad Almisnid
MBBS, MSc MedEd
PhD Candidate,
Leeds Institute of Medical Education (LIME)
Faculty of Medicine & Health, University of Leeds
Work address: 9.12 Worsley Building
University of Leeds, Leeds, LS2 9JT
+966505994439 / +447366319973
umkma@leeds.ac.uk

Supervisor information:
Dr Matthew Homer
Work address:
7.09 Worsley Building, University of Leeds,
Clarendon Way, Leeds, LS2 9JT
Email address:
M.S.Homer@education.leeds.ac.uk

Appendix 13. Information Sheet.

This is the information sheet for the interviews' participants; I also created slightly modified information sheets for the FGs and document analysis.

Leeds Institute of Medical Education
School of Medicine
Faculty of Medicine and Health

UNIVERSITY OF LEEDS

Version 3, Date 21 January 2020

Study Title: Exploring Implementation of the Objective Structured Clinical Examination (OSCE) in Saudi Medical Schools

Participant Information Sheet

Semi-structured Interview

I would like to invite you to take part in a focus group to discuss your experiences with the Objective Structured Clinical Examination (OSCE). Before you decide to participate, it is important that you understand why this research is being done and what it will involve. Please carefully read the following information and discuss it with others if you wish. Feel free to ask me about anything that is not clear and to request more information. Take the time to decide whether you wish to take part.

Why am I holding these interviews?

I am a PhD student conducting research for a degree in medical education at the University of Leeds. I am seeking to understand the OSCE implementation process at your medical school and the challenges and opportunities offered by its adoption. I, therefore, would like to hear about your experiences and thoughts on the OSCE implementation from your position in medical school.

Why are you being asked to take part?

You have been invited to participate in my research because of your role within your medical school. I value and wish to learn about your experiences regarding the OSCE implementation, and your participation in this interview will help me understand how the OSCE is implemented in your context. Anything you share with me will be important and help me understand how the OSCE is conducted in your medical school.

What do I have to do?

You will be asked to participate in an interview, which will consist of a discussion with the researcher. Some open-ended questions about the OSCE will be asked in a way to encourage the discussion. I am interested in listening to your experiences of conducting the OSCE at your medical school, so there are no right or wrong answers. The interview will be recorded privately, and it may last about 40–60 minutes.

If you accept the invitation to participate in this study, please read the information sheet, and sign the consent form I have emailed to you. We will work together to choose the date, time and place to conduct the interview.

With your permission, the interview discussion will be audio recorded as I need to re-listen to and transcribe the discussion to reflect on and analyse it. Only I will have access to the audio recordings,

Appendix 13 continues.

and I will not share them with your school or any other party. Quotes from the conversations may be used in the written report but will not be attributed to any named individuals. I will give you a numerical identifier to ensure your confidentiality, however full anonymity cannot be guaranteed. Because it is important to contextualisation data for analysis and discussion, your role will be identified as academic leader but no mention for specific position (e.g., dean, vice-dean, medical educationist or head of the assessment unit). The school's name will be identified with numerical identifier to preserve anonymity when reporting in the thesis and papers.

All the information shared within the interview will be securely stored at the University of Leeds, and only the supervisory team and me (the researcher) will have access to it.

During the interview, you may take breaks to rest, and you may withdraw at any time. However, due to the nature of data collection process in this research you only allowed to withdraw within the two days after the interview where analysis will begin thereafter.

Will I be paid for taking part?

There is no payment. Participation in this study is fully voluntary but is beneficial as your experiences will be heard and considered.

Do I have to take part?

It is up to you to decide whether to participate. If you decide to participate, you may keep this information sheet and will be asked to sign a consent form. You may still withdraw from the study up to two days after the interview without giving an explanation and without affecting any benefits to which you are entitled in any way.

What will happen to the information collected during the interview?

If you consent to take part, everything you say will be kept anonymous, and the information collected on you will be handled strictly in accordance with the consent you have given and the Data Protection Act 2018 (DPA) https://dataprotection.leeds.ac.uk/wp-content/uploads/sites/48/2019/02/Research-Privacy-Notice.pdf. Your name and contact information will be included only on your consent form, which will be stored in the researcher's secure locker at the University of Leeds. You will be identified with a pseudonym in the transcripts.

To ensure confidentiality, identifiable data will be stored separately from non-identifiable data. The anonymised data related to this PhD study will be retained for three years after submission of the PhD thesis for examination, but no raw (non-anonymised) data will be transferred to the University of Leeds Research Data Repository.

Appendix 13 continues.

Leeds Institute of Medical Education
School of Medicine
Faculty of Medicine and Health

UNIVERSITY OF LEEDS

Version 3, Date 21 January 2020

What will happen to the study findings?

Completing the data collection will take a minimum of three months. Afterwards, a report will be written based on the findings. If you wish, you may email me at any time to request a copy of the report to read for yourself. I will separately publish the study findings in professional and academic journals.

Who is organising and funding this research?

The interviews will be organised and facilitated by Khaled Almisnid, a PhD student at the Leeds Institute of Medical Education. You can contact Khaled for anything related to the interviews. The research is supported by funds from Qassim University, Saudi Arabia.

Who has reviewed the study?

The study has been reviewed by the Faculty of Medicine and Health Research Ethics Committee, University of Leeds, [reference number MREC 18-103].

What do I do if I have concerns?

Please discuss your concerns with the researcher and his supervisors to ensure that they are addressed. Contact details can be found as follows.

If you want further information about the interviews, please contact:

Researcher Name: Khaled Almisnid
Work address: 9.12 Worsley Building, University of Leeds, Clarendon Way, Leeds, LS2 9JT
Contact number: +966505994439 / +447366319973
Email address: umkma@leeds.ac.uk

Supervisor Name: Dr Matthew Homer
Work address: 7.09 Worsley Building, University of Leeds, Clarendon Way, Leeds, LS2 9JT
Email address: M.S.Homer@education.leeds.ac.uk

Co-supervisor Name: Prof Trudie Roberts
Work address: 7.09 Worsley Building, University of Leeds, Clarendon Way, Leeds, LS2 9JT
Email address: t.e.roberts@leeds.ac.uk

Co-supervisor Name: Prof Richard Fuller
Work address: Cedar House, Ashton St, Liverpool L69 3GE
Email address: Richard.Fuller@liverpool.ac.uk

If you decide to take part, you will be given a copy of this information sheet and a signed consent form to keep.

Thank you for reading this information.

Appendix 14. Focus Group Consent Form.

I also designed a slightly modified consent form for the interview participants.

Leeds Institute of Medical Education
School of Medicine
Faculty of Medicine and Health

Version 2, Date 20 November 2019

UNIVERSITY OF LEEDS

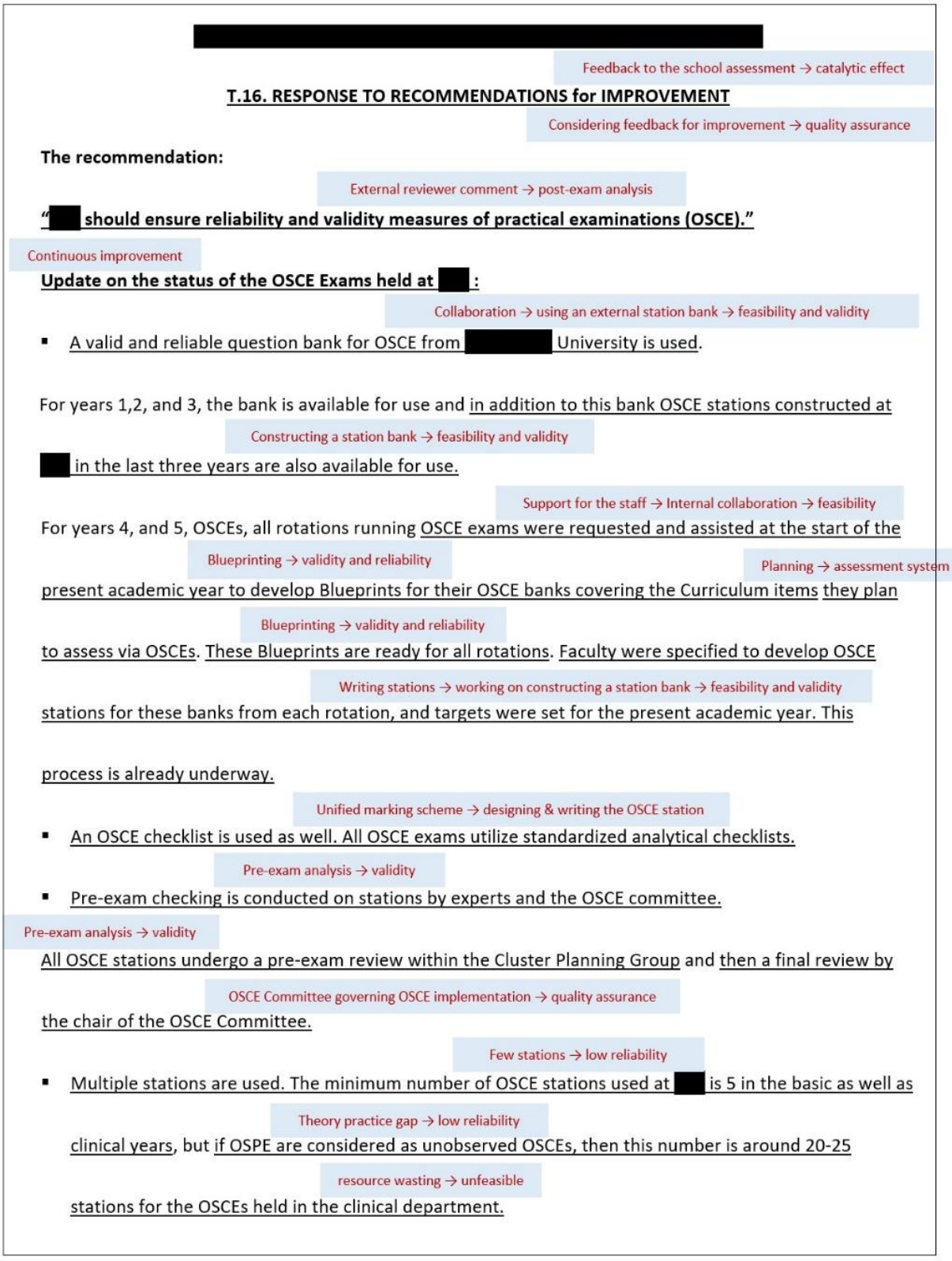
Consent Form

Focus Group

Study Title: Exploring Implementation of the Objective Structured Clinical Examination (OSCE) in Saudi Medical Schools		Add your initials next to the statement if you agree
1	I confirm that I have read and understood the information sheet provided (dated 10 July 2019) that explaining the above research study.	
2	I have had the opportunity to ask questions about the study.	
3	I understand that my participation is voluntary.	
4	I understand my right to withdraw at any time without giving any reason and without there being any negative consequences. However, any data/responses already provided will be retained, as the nature of a focus group makes it impossible to extract the data of a single participant from the group responses.	
5	I understand my right to not answer any particular question(s).	
6	I understand that my name will not be linked with the research materials and that I will remain anonymous. I give permission for members of the research team to have access to my anonymised responses, only if they preserve the confidentiality of the data.	
7	I understand that the personal and identifiable data will be stored in a secure location and destroyed at the end of the study.	
8	I agree to the focus group being audio recorded.	
9	I agree for anonymised quotes to be used in publications.	
10	I agree to respect the confidentiality of the information discussed within the focus group and will not talk to others about what I have heard.	
11	I understand that as part of a focus group, confidentiality cannot be promised on behalf of other participants, although it will be requested.	

Name of participant			
Participant's signature		Date	
Researcher information	Khaled Almisnid Work address: 9.12 Worsley Building, University of Leeds, Clarendon Way, Leeds, LS2 9JT Contacts: +966505994439 / +447366319973, umkma@leeds.ac.uk		
Supervisor information	Dr Matthew Homer Work address: 7.09 Worsley Building, University of Leeds, Clarendon Way, Leeds, LS2 9JT Email address: M.S.Homer@education.leeds.ac.uk		
Researcher's signature		Date	

Appendix 15. A demonstration of the initial code-generating process by using codebook analysis.



[redacted]

Feedback to the school assessment → catalytic effect

T.16. RESPONSE TO RECOMMENDATIONS for IMPROVEMENT

Considering feedback for improvement → quality assurance

The recommendation:

External reviewer comment → post-exam analysis

"[redacted] should ensure reliability and validity measures of practical examinations (OSCE)."

Continuous improvement

Update on the status of the OSCE Exams held at [redacted]:

Collaboration → using an external station bank → feasibility and validity

- A valid and reliable question bank for OSCE from [redacted] University is used.

For years 1,2, and 3, the bank is available for use and in addition to this bank OSCE stations constructed at

Constructing a station bank → feasibility and validity

[redacted] in the last three years are also available for use.

Support for the staff → Internal collaboration → feasibility

For years 4, and 5, OSCEs, all rotations running OSCE exams were requested and assisted at the start of the

Blueprinting → validity and reliability

Planning → assessment system

present academic year to develop Blueprints for their OSCE banks covering the Curriculum items they plan

Blueprinting → validity and reliability

to assess via OSCEs. These Blueprints are ready for all rotations. Faculty were specified to develop OSCE

Writing stations → working on constructing a station bank → feasibility and validity

stations for these banks from each rotation, and targets were set for the present academic year. This

process is already underway.

Unified marking scheme → designing & writing the OSCE station

- An OSCE checklist is used as well. All OSCE exams utilize standardized analytical checklists.

Pre-exam analysis → validity

- Pre-exam checking is conducted on stations by experts and the OSCE committee.

Pre-exam analysis → validity

All OSCE stations undergo a pre-exam review within the Cluster Planning Group and then a final review by

OSCE Committee governing OSCE implementation → quality assurance

the chair of the OSCE Committee.

Few stations → low reliability

- Multiple stations are used. The minimum number of OSCE stations used at [redacted] is 5 in the basic as well as

Theory practice gap → low reliability

clinical years, but if OSPE are considered as unobserved OSCEs, then this number is around 20-25

resource wasting → unfeasible

stations for the OSCEs held in the clinical department.

Appendix 16. An illustration of one key area labelled "criteria for good assessment" developed using codebook analysis.

Key area	Refined codes	Codes
Criteria of good assessment	Validity	Curriculum blueprinting
		Pre-exam analysis
		Post-exam analysis
		Content experts' revision
		Quality assurance
	Reliability	Number of the stations
		Length of the stations
		Managing parallel circuits
		Psychometric analyses
		Standard-setting
		Correlation with other exams
		Marking scheme
	Fairness	Exam security
		Examiner training
		SPs training
		Controlling design
		Oversight the exam
		OSCE design is similar to iterative courses
	Educational impact	Feedback to the students
		Formative exams
	Catalytic effect	Feedback to the school assessment system
		Feedback to the school teaching system
		Continuous improvement
	Feasibility	Station bank
		Practicality
		Availability of internal support
		Receiving external support
		Resources availability
	Acceptability	Stakeholder perception
		No objection from the stakeholders
	Equivalence	Iterative cycles lead to the same decisions

Appendix 17. A comparison of case studies A and B's documents analysis codes with those from the literature.

Themes	Codes from literature Ottawa conference consensus + ASPIRE criteria + WFME standards	Codes from case study A The assessment documents of the medical school A	Codes from case study B The assessment documents of the medical school B
Assessment policy	Assessment policy should be responsive to the context	Yes	Yes
	Specified educational outcomes	Lack of evidence	Mentioned
	Designing & conducting the OSCE	Available but not comprehensive	Available but not comprehensive
	Appeal policy	Available	Available
	Remediation process	Available	Available
Criteria of good assessment	Validity	Indicated with few details	Indicated with few details
	Reliability	Indicated with few details	Indicated with few details
	Fairness	Mentioned	Mentioned
	Educational effect	Mentioned	Mentioned
	Catalytic effect	Mentioned	Mentioned
	Feasibility	Mentioned	Mentioned
	Acceptability	Mentioned	lack of evidence
	Equivalence	Lack of evidence	Mentioned
	Authenticity	Mentioned	lack of evidence
Use a system of assessment	Comprehensive	Mentioned	Lack of evidence
	Coherent	Lack of evidence	Lack of evidence
	Continuous	Lack of evidence	Lack of evidence
	Purpose-driven	Mentioned	Mentioned
	Transparent	Indicated with few details	Indicated with few details
	Acceptable and appropriate	Lack of evidence	Mentioned
	Feasible	Lack of evidence	Lack of evidence
	Quality control	Indicated with few details	Indicated with few details
Educational resources for assessment	Human resources	The lack of staff is mentioned	External examiners are mentioned
	Stakeholder needs	Indicated with few details	Indicated with few details
	Physical resources	The lack of some physical resources is mentioned	The lack of some physical resources is mentioned
	Fund	Lack of evidence	Lack of evidence

Appendix 18. Case study A documents: Excerpts from the medical school A Assessment and Evaluation Guide.

Purpose of assessment:

Purposes of assessment include:

1- **For learner:**

- Motivation and support of student learning.
- Diagnosing student weakness and strength.
- Measuring achievement of objectives.
- Testing progress and improvement.
- Measuring competence.
- Self-evaluation.

2- **For institution:**

- Showing curriculum effectiveness.
  - Effective teaching.
  - Achievement of objectives.
- Quality assurance.
- Motivation of teachers.
- Predicting future performance.
- Introducing curriculum change.
- Certification and judgment of competency.
- Development and evaluation of teaching programs.
- Understanding of the learning process and student progress.
- Ranking of students and promotion from grade to grade.

Criteria of good assessment:

- **Fair**
  - Fairness in exam content and grading system.
- **Valid**
  - Tests what needs to be tested.
- **Reliable**
  - Results are almost repeated.
- **Motivating**
  - Assessment drives learning.

Objective Structured Clinical Examination (OSCE):

Description:

OSCE is an assessment tool in which the components of clinical competence such as history taking, physical examination, simple procedures, patient management problems, communication, and attitude are tested using agreed checklists (a standardized marking scheme specific for each case is used.) and rotating the students round a number of stations some of which have observers with check lists.

OSCE is applicable to any situation where clinical competencies are to be tested. It consists of multiple stations (usually 15-20, each station takes around 15 minutes), where each candidate is asked to perform a defined task as mentioned above.

OSCE stations, some of it might be silent; consist of skill of reading an EKG, Chest X-ray, Laboratory abnormality and/or a picture of abnormal signs etc. Number of silent can be up to 7 stations out of 15 and the remaining are pure clinical examination.

Advantages:

- An effective alternative to unstructured short cases
- Allows wider sampling and standardization of cases
- Greater reliability of marking

Objective Structured Practical Examination (OSPE):

Description:

OSPE is a new concept in practical assessment of basic medical sciences; it is a tool which is used to assess laboratory skills (e.g. anatomy) of students in the preclinical stage of a medical curriculum. It is more objective, reliable and a valid tool of assessment to assess practical aspects in integrated courses.

In OSPE, multiple stations (usually 13-20, each station takes around 2 minutes) are designed and each station has a specific objective which needs to be tested.

Advantages:

- Assess higher cognitive skills like clinical application.
- It is objective, reliable, valid and discriminatory
- All students are exposed to same standardized questions
- It covers a wide spectrum of learning domains
- It tests a wide range of skills in a short period of time

Appendix 18 continues.

The Use of Rubric in Assessment of Skills and Knowledge:

Often, rubrics are associated with skills assessment. However, rubrics are not just for skills, projects, or processes. Rubrics are used for assessing knowledge, writing essays, or the aspects of written assignments.
For example, senior medical students are asked to evaluate the impact or problem an issue has on society. The progression through the rubric goes from lower-order thinking skills to higher-order thinking skills (Bloom's taxonomy):

- **Recall:** outline, list, define.
- **Comprehension:** describe , explain.
- **Application:** interpret, demonstrate.
- **analyze:** compare, illustrate.
- **Synthesis:** arrange.
- **evaluate:** justify, appraise.

Steps for Development of rubric:

Stage 1: is for developing rubrics if the task is not already defined.

1. What are the learning outcomes?

The first question to ask is, "What are the learning outcomes and Why are we doing this assessment?" If we don't need to do an assessment, then why do it?

2. Identify the key elements or components:

Once the learning outcomes for the assessment have been established, the aspects of the assessment that we are going to assess need to be identified. Are we assessing one aspect of the project development process? Are we addressing other aspects instead, such as knowledge, process, and so on?

Listed here are the 4Ds:

1- define.
2- design.
3- do.
4- debrief.

3. Clearly define the task:

The next three parts of task development are intimately linked together. Here we define the assessment tasks that are linked to the project the students will undertake.
Assessment tasks are best when they are straightforward and uncomplicated.

A structure to consider when writing tasks and developing projects is *SMART*.
This is an excellent guideline in an easy-to-remember form, and it covers off the key points that define engaging and relevant tasks. When applying SMART, determine if the task is:

- Specific.
- Measurable.
- Attainable.
- Realistic.
- Timely.

It is best to ask, "Is this a SMART task or goal?"

4. Design the task:

Again, we need to ask key questions:
Is it suitable for the purpose?
Does it link to the curriculum?
Does it link to learning objective?
Is it SMART?
Is it suitable for our student?
Can students this level achieve at the highest level?

5. What are the expected outcomes?

When we set a task, we need to consider what it is we want to have our students achieve. The best results come when you have an exemplar of the outcomes you expect. If this is not available, **clearly define the standards:**

- Excellent standard.
- Good standard.
- Acceptable standard.
- Poor standard.
- Failing.

Appendix 19. Case study A documents: Excerpts from the medical school A Assessment Policy.

Article 14

Objective Specific Clinical Examination (OSCE)

Section A: definition and description.

14.1. Objective Specific Clinical Examination (OSCE) is an assessment tool in which the components of clinical competence such as history taking, physical examination, simple procedures, patient management problems, communication, and attitude are tested using agreed checklists.

14.2. An OSCE station may have a written component to be done by a student at the end of it. This written component is directly relevant to the OSCE and usually consists of a note to be filled or questions regarding management and disposition.

14.3. A group of OSCE stations might form an assessment scenario.

14.4. The content of each station might be a standardized patient, simulated patient, another health care worker, variable fidelity manikins or a layperson.

Section B: Conduct.

14.5. All OSCE stations are video recoded and remotely monitored and graded by evaluators.

14.6. Interaction and communication between a student and an evaluator is discouraged unless extremely needed.

14.7. If a manikin with events log capability is utilized, a copy of the log is kept with the scoring checklists for reference.

14.8. Every OSCE station should have an instructional sign indicating the required task to be performed, the sign should be short and clear.

14.9. After every OSCE station, students will have a rest of an equal time to the time allocated to the stop, which is utilized by technicians to reset equipment.

14.10. Evaluators are not allowed to be inside the room of the OSCEs, they should instead watch the student's performance using audiovisual systems.

14.11. The exam will be taped including reset times and including the hallway where students stay during the reset time.

14.12. Tapes will be kept for a maximum of ten days securely in a server that has high security barriers.

Section C: Scoring and key answers.

14.13. An OSCE station may have more than one evaluator and the score will be calculated based on average.

14.14. A detailed action card of every OSCE station shall be submitted as a key answer; it shall have all possible correct answers and detailed scoring scheme.

14.15. Every OSPE station should have a preformed rubric detailing the scoring scheme.

14.16. Course organizers should submit a contingency plan for equipment failures.

14.17. The "correct" result of a test or the correct diagnosis in an OSCE station shall have minimum scoring not exceeding 5% of the score of the stop. It is recommended not to score results, as the outcome intended is the technique itself.

14.18. The pass/fail cut off for each OSCE/SOE station is determined by the course organizer or clerkship director prior to conducting the exam using a Minimum Performance Level (MPL) scoring system.

14.19. Each station shall be assigned an MPL based on the expected performance of a minimally competent student. The assessment unit shall approve station MPLs prior to the exam.

Appendix 19 continues.

Article 23

Development of assessment questions

Section A: written assessment.

23.1. Assessment item writers shall follow the Assessment Unit guidelines on developing written assessment tools.
23.2. Assessment item writers shall have formal training sessions on tools they wish to use.
23.3. All new questions written by faculty content experts are initially edited by the course organizer or clerkship director and returned to the faculty content expert for approval.
23.4. After course organizer or clerkship director and faculty content expert(s) have completed their internal review of new MCQs, the new questions are submitted to the assessment unit for final approval.

Section B: observation based assessment.

23.5. Assessment item writers shall follow the Assessment Unit guidelines on developing observation based assessment tools.
23.6. Assessment item writers shall have formal training sessions on tools they wish to use.
23.7. For OSCE assessments, the course organizer or clerkship director shall fallow Angoff's method in development.

Article 24

Post hoc analysis and interpretation of results (for high stakes scenarios)

Section A: Item Analysis Report (IAR).

The Assessment Unit will provide course organizers and clerkship directors with a report entitled Item Analysis Report (IAR). This document is prepared to help instructors at [redacted] interpret the statistics generated about their assessment sessions and improve the effectiveness of test items and the validity of assessment scores. The IAR shall consist of:

A. Correct responses as a percentage of the total group.
B. The difficulty index of the assessment session.
C. Correct responses as a percentage of the upper/lower 27% of group.
D. Discrimination Index (Point Biserial).
E. The standard deviation of scores; mean and median scores.
F. Nonfunctioning distractors if applicable.
G. Percentile scores for all students.
H. Reliability coefficient (KR20).
I. Suggestions on improving the score accuracy.

Section B: IAR interpretation and usage.

The main goal of generating an IAR is to ensure validity and reliability of assessment tools. The assessment unit shall keep a master record containing IARs for all courses and clerkships. An annual report on IARs may be generated upon the request of the vice dean for academic affairs.

Another usage of the IARs is to correct for errors and illuminate unusually difficult questions from the assessment final score. The assessment unit shall provide course organizers and clerkship directors with suggestion at the end of the IAR on how to correct for such errors.

Appendix 20. Case study A documents: Examples of marking schemes used in two different courses at medical school A.

MD 3 – SURGERY: Abdominal examination station

Students name: ..... ID: .....

	Skill	Mark	Done	Not done
	1-Great the patient, introduce self, take permission.	1		
	2- Inspection: From the foot end and the Right side	1		
	Comment on: contour, umbilicus, hair distribution, veins, scars, masses , renal angles , spine , hernial orifices	1		
	3- Palpation: superficial palpation start from proper site and watch for facial expression	2		
	Measurement of Liver span / palpate Spleen /palpate kidney	1		
	Elicit Murphy's sign	1		
	4- Percussion: Traube's area / Shifting dullness	1		
	5- Auscultation: intestinal sounds / bruit	1		
	6- at Conclusion no to forget: Mention DRE and hernia orifices Thank the patient.	1		
	7- Discussion: DD of right upper quadrant abdomen pain	2		
	What are the signs of acute cholecystitis?	2		
	What are the complications of acute cholecystitis?	2		
	What are the investigations in evaluation of acute RUQ pain?	2		
	What are the management options for acute cholecystitis?	2		

Total Marks: ...../ 20

Final Clinical Exam - Orthopedic Surgery Course

Student's name: ..... **University ID:** .....

Section 1		
History 1. Chief complain and duration 2. History of presenting illness A. History of pain (SOCRATES) 3. PMH 4. PSH 5. Allergy and meds 6. Social history 7. Family hx 8. Systemic review		**4**
Investigation 1. LABs (CBC, ESR, CRP) 2. Joint aspiration 3. X-ray, MRI, Bone scan		**3**
DDx: septic Arthritis, OM **Treatment** **1.** Surgery		**2**
Total		**10**

Appendix 21. Case study B documents: Excerpts from the medical school B Exams Policy.

EXAMS POLICY

Index

EXAMS POLICY

Introduction:

- The pre-exam analysis meeting is a **REVIEW PROCESS** in which the exam review committee implement the quality assurance in the exam paper. It is **NOT a PREPARATION process,** in which the exam paper is generated from the start. This concept should be in mind of all staff members.
- The submitted paper must be the finalized version (after agreement between HOD, coordinator and tutors and completed final editing).
- Taking the time to describe errors on an exam and to identify the outside factors contributing to performance is a useful strategy for improving future exam performance.
- Referring to the yearly Academic Calendar officially published by the Vice Dean's Office, all coordinators should integrate all pre and post exam analysis meetings into their timetables from the beginning of each block/cluster/rotation.

Pre-exam analysis

Calling for pre-exam meeting:

- **Venue:** Assessment unit (AU)
- **Timing:** Only in any of the following time slots:
  - **BASIC Exam Review Committee:**
    - Sunday: 1:00 – 3:00 pm.
    - Wednesday: 1:00 – 3:00 pm.
  - **CLINICAL Exam Review Committee:**
    - Monday: 1:00 – 2:00 pm.
    - Tuesday: 10:00 am – 12:00 pm.
    - Wednesday: 11:00 am – 12:00 pm.
  - Any of the two departments (i.e. Basic and Clinical) can hold their meetings in the other's slots (when the time slot is free) ONLY in emergency cases with approval of respective head of department and head of assessment unit.

Appendix 21 continues.

Exam Analysis Policy

The HPEU analyzes exam data according to the following criteria:

- Very easy: ≥ 95%
- Easy: 90% - 94%
- Difficult: 10% - 20%
- Very difficult: ≤ 10%
- Discrimination: below zero (negative value)

General guidelines:

❖ **Facility index**

- Easy questions (it should not exceed 25 %)
- Difficult Questions (it should not exceed 25%) with justification.
  - If the easy & difficult questions exceed the allowed percentages mentioned above, the HPEU will deal with each case individually based on the type of exam, purpose, level of learning objectives and KPIs.
  - The HPEU will consider all of the following when reviewing cases:
    - Students' feedback for the block/exam
    - Tutors' reply to feedback/justification
    - Pre-exam analysis quality
    - Blueprint alignment

Based on the above guidelines, the HPEU may delete some questions, keep the exam as it is, recommend improvement plans and/or do further investigation.

❖ **Discrimination index**

- Negative value questions are to be reviewed and/or eliminated.
- The review process starts with the Coordinator and then transitions to the Post-Exam Review Committee.

❖ **Coefficient of internal consistency (Reliability Coefficient):**

The following general guidelines determine the reliability coefficients for SRC exams:

Reliability	Interpretation
≥ 90%	Excellent exam reliability up to the highest standard levels.
89% - 80%	A very good exam reliability.
79% - 70%	A good exam upon average. However, some items could be improved.
69% - 60%	A somewhat low exam reliability. A substantial number of items could be improved.
59% - 50%	A low reliability exam which needs revision or inclusion of more exam items.
≤ 50%	The reliability of the exam is in doubt. Major revisions are needed to the exam and its items.

- The HPEU will take into consideration several factor related to exam reliability (for example; number of questions and number of students).

Appendix 22. Case study B documents: Excerpts from the medical school B template for an OSCE station.

OSCE Examination
OSCE CVS [redacted] in OSCE bank. Year: 4 Rotation: A Station System: Cardiology Date: 4 Apr 2021

Examiner Guide

Instructions for the Examiner:

Assignment 1&2

The list of criteria applies a four-point scale. The instructions for the observer provide a general description of the actions the student should perform. You have to indicate whether the student's performance is Good, Moderate, Poor or Not carried out at all.

This description of the OSCE provides information about the assessment of items marked with * in the score list.

Assignment 3

The manikin has a Pan systolic murmur in the left 5th intercostal space 2 cm lateral to the Mid Clavicular line. The murmur radiates into the Axilla.

Assignment 4: Questions to ask the candidate:

1. What are the auscultatory findings?
   Pan Systolic Murmur at the cardiac apex.

2. What is the diagnosis?
   MR.

3. Give three causes of Mitral Regurgitation.
- Mitral valve prolapse.
- Damaged tissue cords.
- Rheumatic fever. Rheumatic fever
- Endocarditis.
- Myocardial Infarction.
- Abnormality of the heart muscle (cardiomyopathy).

4. What are 3 complications of MR?
- Heart failure.
- Atrial fibrillation.
- Pulmonary hypertension.
- Thromboembolism

5. How is this valve disease managed?
- Cardiac dilatation: Ace Inhibitors
- Angina: Beta Blockers, Calcium Channel Blockers (Nitrates contraindicated)
- Hypertension: Ace inhibitors/ Beta Blockers, Ca Channel Blockers
- CHF: Digoxin/Diuretics
- Surgery: Valve replacement

OSCE Examination
OSCE CVS [redacted] in OSCE bank. Year: 4 Rotation: A Station System: Cardiology Date: 4 Apr 2021

Simulator Guide

If the OSCE station incorporates a Patient/patient simulator this section should provide a detailed guide to the simulator on the OSCE station as well as instructions on the simulation of history/symptoms to be performed by him/her.

This should be discussed with the simulator and the simulator prepared comprehensively preferably prior to the OSCE exam day itself.

A mock run should be arranged with the simulator as well.

Appendix 22 continues.

OSCE Examination
OSCE CVS [redacted] in OSCE bank. Year: 4 Rotation: A Station System: Cardiology Date: 4 Apr 2021

Guide to marking the checklist

	Steps of the Skills to be demonstrated **Assignment No 1 & 2:**	Good (3)	Moderate (2)	Poorly done (1)	Step missed (0)
1	Proper position of the patient in bed at 45° and contralateral rotation of the neck	Performs the skill effectively and fluently	Perform the skill ineffectively with some lack of technique	Performs the skill with grossly wrong technique	
2	Put a horizontal ruler at the upper level of Rt. jugular vein and put a vertical ruler at the angle of Lewis				
3	How to calculate JVP in cmH_2O				
4	Inspection of the precordium				
5	Palpation of the apex				
6	Palpation of the Tricuspid area/A2/P2 and parasternal heave				
7	Auscultation of the apex with the diaphragm				
9	Timing of S1 with Carotid palpation				
10	Auscultation of the ERB point				
11	Auscultation of the Pulmonary area and the aortic area				
	Auscultation of lung bases				
Assignment No 2: (Auscultates the manikin for 2 minutes)					
Assignment No 3:					
12	Describe your auscultatory findings	Correctly describes the pansystolic murmur	Picks the murmur but is not clear whether the murmur is Systolic or diastolic	Auscultated the murmur but is not able to describe it	Misses the Murmur completely
13	What is your diagnosis	MR		Wrong diagnose	No diagnosis
14	Name 3 complications of MR	3	2	1	0
15	How is it managed?	States 3 options	Mentions 2	Mentions 1	Does not any

Examiner name	Final grade	Score /10	Remarks

OSCE Examination
OSCE CVS [redacted] in OSCE bank. Year: 4 Rotation: A Station System: Cardiology Date: 4 Apr 2021

Student Assignments

Dear Student,

You are in OPC; your 75-year-old patient has presented with progressive shortness of breath for the last month, he had 3 coronary artery stents implanted following an acute coronary attack two years ago.

Assignment no 1 (4 minutes)
Demonstrate how to perform clinical measurement of the jugular venous pressure on the simulated patient.

Assignment no 2 (2 minutes)
Perform precordial Examination
NB1 Mention the rationale for each auscultatory area while you are examining.
NB 2 Ensure positioning of the patient as per required for each valve disease.

Assignment no 3 (2 minutes)
Auscultate the heart apex of the manikin.

Assignment no 4 (2 minutes)
After the examination, the examiner will ask you some questions.

Good Luck!

Appendix 23. Case study B documents: Excerpts from the medical school B OSCE pre-exam analysis review (left), and an external examiner's post-exam feedback (right).

OSCE Stations Pre-Exam Review Family Medicine Rotation A

Year 5

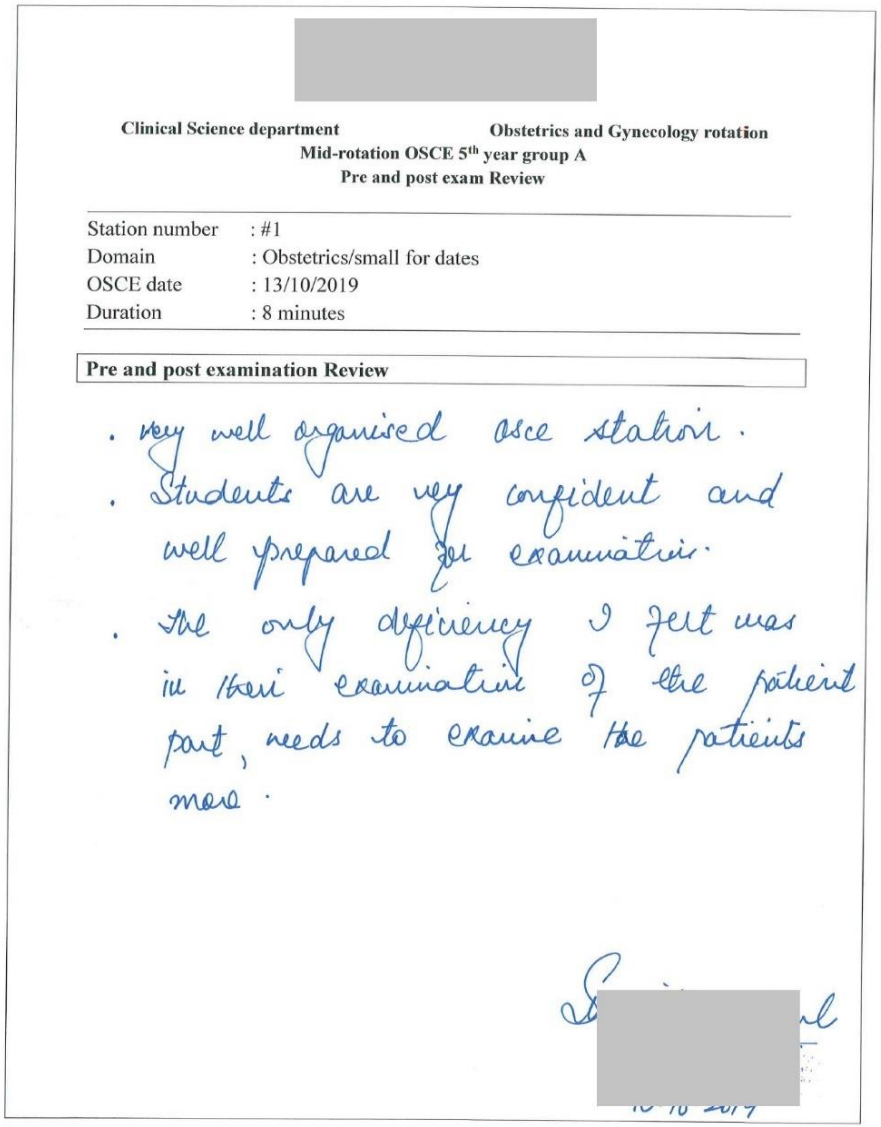
Academic Year 2019-2020

Case: Asthma

1. Please add "global score" to the scoring checklist allowing a more subjective reflection of the examiner on the student performance as well.
2. Please add the "Detailed Examiner instructions" as the case may be evaluated by more than one examiner and this will help in calibration of the station as well more uniformity in the scoring.
3. Please clearly indicate the total time and the time allotted to each assignment within the station so that the student can be guided not to take too much time on one assignment and to also finish the station within the prescribed time.
4. Please add the "list of requirements" for the station so the required logistics for the station could be arranged.
5. Please use the detailed "OSCE template document" that has been finalized for all the clinical department OSCE exams.
6. The OSCE station needs to be labeled in accordance with the ILO so that it can be retrieved later form the OSCE station bank.

Case: Geriatric Examination

1. Please add "global score" to the scoring checklist allowing a more subjective reflection of the examiner on the student performance as well.
2. Please add the "Detailed Examiner instructions" as the case may be evaluated by more than one examiner and this will help in calibration of the station as well more uniformity in the scoring.
3. Please add the "list of requirements" for the station so the required logistics for the station could be arranged.
4. Please use the detailed "OSCE template document" that has been finalized for all the clinical department OSCE exams.
5. The OSCE station needs to be labeled in accordance with the ILO so that it can be retrieved later form the OSCE station bank.



[redacted]

Clinical Science department **Obstetrics and Gynecology rotation**

Mid-rotation OSCE 5th year group A

Pre and post exam Review

Station number	: #1
Domain	: Obstetrics/small for dates
OSCE date	: 13/10/2019
Duration	: 8 minutes

Pre and post examination Review

. very well organised osce station.
. Students are very confident and well prepared for examination.
. the only deficiency I felt was in their examination of the patient part, needs to examine the patients more.

[redacted]

Appendix 24. An excerpt from an interview with an assessment specialist.

The researcher:
Good. so let me move to another point. What are the major advantages and disadvantages of the OSCE?
The participant:
Yeah, you know you're never going to have a perfect assessment. The perfect assessment tool does not exist, right? We have to be very clear about it. Right. First of all, I mean it's I mean we talked about the advantages. It allows you to test skills on students that otherwise, it would allow, you know, practising on patients in a safe, controlled environment. Right? So, we have partly trained students that are going through the skills training. We do not expose patients to harm and do not, you know, so it gives us a very, you know, useful, you know, of course, the issue that we can test many types of skills in it. So, it is not just procedures; even communication skills can be tested. It's a tool that can assess a lot of types of variables, you know, breaking bad news communication and use it to test CPR, something as a high stake as CPR. So, it is a very flexible tool. So, it's got a lot of variation, and it's very flexible, so that's another advantage. You could, I mean if you look at the structure of the exam, you can assess when it's designed, you can assess a large number of students in the same standard environment. Yes. So, we have the same students going through the same OSCE cycles with the same exam and the same simulator. So, it tends to, let's say, be quite relatively fair for the students we consider it.
It has its downs. It is a very time-consuming assessment to design. Implementing it does take a lot of effort. A lot of resource goes into implementing and designing it. That is one issue time intensive. You could say that the advantage that I mentioned is that it's a safe environment, but people who proponent voting against it would say that OK, you are, I mean, considering it as a safe environment, but it's not the real environment. So, they could say, OK, it could be, you know, the value of the assessment can be questioned in that. Basically, the student knows it's an exam, so they're under the pressure of the exam or a student who's under the pressure of trying to resuscitate a patient in real life. The performance might be quite different.
The researcher:
Yeah.
The participant:
If we consider it, yes. And then, if we considered it as a summative tool, you could have students who had a bad day, or they could have had a bad case that doesn't truly represent the skills of the student; I mean, you could have a bad or a good day. When you use it as a single summative tool, then it might not be truly reflective of the students, you know, skills overall in particular, you know, curriculum or particular syllabus. So, it has its negatives, and it has its positives.

The researcher:
All right. Very nice. And so, let me move into your college now, and you know OSCEs have multiple steps to perform and implement. So, can you please give me a quick sequence for how your OSCE is conducted in your college?
The participant:
You know it's a lot like you know a well-prepared food, let's say a lovely dish it's eaten out in minutes. Yes. But the time it takes to prepare it that's quite long; OSCE is exactly like that. I mean, a typical OSCE plan would start off at the start of the academic year. So, we would look at the learning we would in the cluster planning group; we would identify the skills and the learning objectives that will be assessed for the true OSCE assessments. That's the way we identify the skills in all courses, and all of them have identified this it's built into the curriculum. Next, a blueprint is looked at, so we take those skills and blueprint them. We make an OSCE assessment blueprint at the start of each rotation. So, the skills go into a blueprint, and once the blueprint is ready, then the faculty, and the tutors, are invited to start preparing and structuring OSCE assessments.
The researcher:
Writing the questions. You mean.
The participant:
Writing the questions, we have basically very comprehensive documents for an OSCE. So, it starts off that day with the learning objective or the OSCE and how it fits into the course objectives and into the program objective. So, it's identified the course learning objectives and the program learning objective that question we'll try to address. Then we, of course, have a next; we structure the assignments for the student that they expected to perform. We then structure the checklists specific to those assignments. After the checklist is designed, we actually then have we have examiner notes. So, each of the steps of the assignment or each step of the checklist is then described for an examiner once that's, so, it's elaborated in detail. So, we have a checklist item. Each step is elaborated on individually in the examiner's notes. We have then a form that says logistic requirements for the OSCE. So, all the requirements, from the simulator to the pencil to the hand sanitiser and the bed sheet, which structured out for the skills lab, so they were able to prepare it.
And once this is done, this is discussed the OSCE design discussed in the planning group. It's the start of the Pre-exam assessment. So, we have the pre-exam let's say, evaluation or the pre-exam review of the OSCE station. And once that takes place, the OSCE, the full document is sent to the OSCE committee for review. So, I'm the chair of that committee, and I have members Mr. (xx), and we have members representatives from the clinical side as well as the basic sciences. And we review the OSCE stations. Primarily what actually happened is I reviewed them individually and wrote my feedback on each station, and then I

sent it to my colleagues for their feedback and suggestions within the OSCE committee. After these suggestions and feedback, this is the pre-exam review; we do send that reviewed OSCE station to the examiner. I could give you examples. The type of review that we do is we might think that the committee might think that the 10 minutes OSCE the steps described out assignments that describe out might not the time is not sufficient. So, we give all the feedback so that finally you know if it's possible to restructure the assignment; we might have been concerned about the time. On the other side, we have, OK, you have 10 minutes, and the assignments we believe would be over in six minutes. So, if it's possible, can you please utilise the next four minutes productively? So, we give that feedback.

And, of course, once the OSCE goes back to the OSCE Committee. They make sure that the relevant suggestions and feedback are incorporated. And then, we have the ingredients ready for the dish. So now we have the OSCE station ready. We then have meetings with the skills lab; the OSCE exam as a whole is sent to the skills lab, and they prepare all the logistics.

So, for the room and everything that will be prepared before the exam, we have a calibration meeting with the examiner. And I think that's probably, one of the most important things we do is we have the same OSCE station. We have multiple tracks. I mean, presently, we have the male college track and the female college track. And if you have a lot of students, for example, in the basic years, we divide them into multiple tracks. So, we would have the same OSCE station being implemented and supervised by two or three different examiners, one in each track. And what we do is we may take a lot of effort into the stations' examiners to put them together, and they calibrate the OSCE stations among themselves. So, they are bringing a scheme on the checklist best that the students will get a reliable a fair assessment from all three examiners. Because we like to achieve high interrater reliability, so, when we come out of the OSCE, we believe that the three stations, the same standard being administered and maintained by three examiners, would have been administered in a uniform way.

The researcher:

So, for each station, you have three assessors!

The participant:

No, we tried to assess one station would usually pertain to one system, but it could try to assess different, you know, for example, OSCE station might start off with a brief history taking and then do an examination, and then maybe we present data to the student who tries to interpret.

The researcher:

But one examiner.

The participant:

Yeah.

The researcher:

All right. So, what after the exams, when the exam is finished?

The participant:

I said that this is an effect that we have worked on, you know, that's the most recent improvement we have been doing the post-exam process. So, we are aware that the post-exam review is very important for any assessment. And we have three aspects of feedback that we take. So, we have student feedback on the OSCE. So, the students receive forms as soon as they leave the exam, and they can anonymously give their feedback on any of the stations actually on any of the examiners on any of the checklists. So, we collect that feedback; that's one aspect of the feedback.

For the second feedback, we usually invite external examiners. So, the external examiners from outside SRU are specifically given a form to write their comments and their feedback on the OSCE assessment. So that's external feedback. And we give a lot of value to that because naturally, it's much more, let's say, objective.

We have a post-OSCE examiner meeting, a short meeting of all the examiners that participated, and we take feedback from them. If they had any difficulties in the logistics, did they pick out any skill that they thought that consistently the students were not performing well at? Did they have any feedback on the checklist items or anything at all? So, we had student feedback. We specifically ask the external examiners to give their written feedback, and we have a meeting with the internal examiners, and we discuss how the whole exercise plan, so any feedback is there.

The researcher:

Yeah. And how about the feedback from examiners to students?

The participant:

Yes. During the OSCE, they're not allowed to give feedback at all.

The researcher:

No written feedback?

The participant:

No. We invite the students for a post-exam meeting. Actually, we have post-exam feedback meetings with the students for each course. And that's not only for the OSCEs, for it's a meeting for when once they've had the final exam and as a part of that post-assessment meeting between the students where the feedback is highlighted, and we do a discussion and take feedback from the students. A part of that is actually one aspect given for OSCE feedback.

So, the students have feedback that they're given in writing, and they want it to be discussed. The examiners want to give them feedback. We have collected the coordinator

might've collected from the different examiners. So that's where the students and the tutor or the exam coordinator will interact, and you know, exchange feedback on the OSCE as well as the other assessments. And it happens at the end of every course.

The researcher:

How do you evaluate your OSCE? For example, do you have any psychometric analysis for your results?

The participant:

Yes. Yes. Since last year, we wanted to do I mean we have taken up this project of actually going through the psychometric assessment, and we are collaborating with Mr. (xx) in health education. We have started off with the very basic psychometric assessment as our first step. So, what we do is that we look at individual stations, and we, for example, we have an OSCE exam that had seven stations, and we actually compare and contrast the scores of the students of the cohort on these seven stations in between. So, we had seven stations, let us say one of the averages for the station was 70%, but for one station at 65% or 60%, the other station had 90.

So, at the moment, when we are testing the comparison, we do a very limited statistical analysis of comparing stations among themselves; that's one aspect. So, we like to highlight, or you know, let's say, pick out any stations in which the students apparently performed poorly. That's one aspect. Then we look at actually if we have two tracks, then we compare the tracks between themselves. So, we look at seven stations here and seven stations there, seven different examiners on each side. The cohort is the same group of students that have been randomised. They are randomly distributed. And we like to compare the individual OSCE stations with each other over the two tracks. That's the second step of, let's say you could say, limited statistical analysis of our OSCE stations.

We do try to indicate if an item in the checklist was picked up by an examiner as being particularly low in performance. For example, students persistently tend to miss one important item on the checklist; then, we invite the examiners to identify it. So, let's say I could give an example of taking consent. So, the examiner notes that the students many of them missed out on this important point that is highlighted. And then, we take that as feedback on the teaching and the learning process as well.

We are looking at the possibility of doing a proper psychometric analysis, including the individual items in the checklist. That is what we want to achieve next, Insha'Allah.

The researcher:

How about reliability? Do you have any calculations for reliability?

The participant:

We calculate the alpha coefficient, which is automatically done for all the exams. So, and I mean, we believe that comparing the two tracks comparing the results with the previous OSCE results adds a bit of reliability, but presently, unless we go into analysing the individual checklist items within each OSCE, I believe that we still have a long way with the psychometric analysis.

The researcher:

Alright, nice. So, let me move a little bit away from your college and then ask how the regulatory bodies how it affects the OSCE and assessments in your college. Do they have any influence on your assessments OSCE assessments?

The participant:

Yes. I think that I like to speak about the SaudiMEDs, and I think that, interestingly, our program actually has been derived from the SaudiMEDs. So, the program learning objectives and the program are basically the strategic plan for SRU. We had a look at the SaudiMEDs, and that program learning objectives were derived from it. So SaudiMEDs incorporates skills, you know, psychomotor skills as one important aspect. And that has been translated into the programme learning objectives. And from there, the programme learning objectives it has actually filtered into each department's course learning objectives and then into the intended learning objectives. So, I like to look at it; I mean, there is a trickle-down effect. So, when we designed the OSCE and blueprinted the learning objective, we had to identify the course learning objectives. And then the program learning objectives; once we identify the PLO, then this is automatically linked to the SaudiMEDs that's where we derive from, and that's for the SaudiMEDs. So, I believe we are definitely aligned with the SaudiMEDs in all assessments.

Now the NCAAA we had the review visit, and we are in the process of the NCAAA accreditation, and we are using the feedback that we had from the NCAAA review visit last year, and they gave us a provisional accreditation. They give feedback on all aspects of the program. And one of the aspects that they gave feedback on was the OSCE, and we, you know, taken those recommendations very seriously. And this year, we've been working on this. I mentioned the psychometric, you know, post-test analysis; it was one thing that they actually identified. So, we are thankful to them for, you know, giving us this feedback and giving us initiatives, we are working on. (NCAAA accreditation visit and feedback)

Now the Ministry of Education they have had regular visits to the college, and they basically have gone through each aspect, like assessments in our college. And I'm directly not, you know, aware of the feedback that we've had from them, any formal feedback. But generally, the feedback for our students that we heard, which can be provided in a documented form when they are going out into practice as interns or graduates regarding skills, is quite high. So, I mean, we tend to feel comfortable regarding our OSCEs, but of course, the room for improvement is there, and we are always continuing to improve our standards.

Appendix 25. An excerpt from a focus group with OSCE designers and examiners.

The researcher:

The next question is: What do you think about the quality and appropriateness of your OSCE overall? You have mentioned some opportunities and challenges, but how do you evaluate your OSCE in general?

Participant 6:

It needs improvement. I mean, we know what the real OSCE is, and we know what we are doing. So, there's a gap between the two. That gap needs to be filled.

Participant 4:

For example, we want to keep ten to twelve stations in a row. Then, students will circulate, the bell will ring every five minutes, and one student will go to another station like that, like the real OSCE, like we do the exam in the Royal College. And when we got the MRCS exam, we rotated; there was a big room, and there were some stations. We rotated, and the bell rang every five minutes we were moving. We took our papers and filled the papers. And it's not that difficult. If we get a big space, if we get logistics, we can do it.

Participant 3:

I agree with what Dr. (x) said. We are making the best of our efforts in this situation.

The researcher:

Alright, very nice. And how about others? Are they pleased with what they're doing right now? or do they have different opinions?

Participant 1:

Of course, we are not happy. It's better to have a real or a good way to assist the students like an OSCE, but lack of facilities and lack of faculties who can assess the students in a good way are big obstacles that we are facing; a lack of staff, lack of facilities, and as ophthalmologists consider a subspecialty, which is a deep one. It needed a sophisticated instrument, sophisticated mannequins, to apply for the OSCE exam. For real patients, it's difficult to get one for simulating one. It's also difficult to get one.

Inshallah, we will try in the future, but I don't think it will be easy to apply for college. It's not like medicine, general surgery, paediatrics, or obstetrics, the main or major specialities in medicine. For those subspecialties, it is considered a bit difficult. But we'll look for a better way to assess our students, inshallah, in the future, so we need to discuss this with our director Dr. (x). Inshallah, we'll try to get our best Inshallah. But now we are not happy, of course.

Yes, I agree; some courses like ophthalmology and ENT still lack the clinical OSCEs and based on what we agreed on as a definition for the OSCE. That is because of the course design and the curriculum itself. That we what we're facing now because we have only two weeks, or less than two weeks, to give them vertical deep knowledge, which they have almost less than maybe ten per cent which that they are familiar with, or they have exposure to it in their previous basic science. But in the new curriculum, we are implementing this soon because the bedsides or the clinical training during the course will be more.

So, we will apply to the OSCE; however, it will be limited because of these difficulties mentioned by Dr. (x) because it needs more skills. So, we're just trying to assess their understanding of what we give them, even during their bedside teaching. But at this moment, we are not happy at this moment we are not applying in ophthalmology or ENT, just applying picture-based scenarios.

The researcher:

OK, anything else before moving on to the next question?

So, do you employ any quality assurance measures like the post-hoc psychometric analysis for your results? Or giving feedback to students or receiving feedback from students about your exam? So, nobody does a psychometric analysis for the results.

Participant 3:

No, we're not doing psychometric analysis or giving feedback to students. However, we take feedback on the whole course, including the assessment way. But again, it's not a well-structured feedback way. Because, how I can say this one, it was in all like a one-man effort to build this up, even though I do not think it was based on Medical Education's evidence for assessment way or feedback way. However, we are not doing psychometric analysis, only just feedback on the course.

Participant 1:

The evaluation has to be from a third party, like a medical education department, who will give the feedback to the course organiser. They will take that feedback from the students, analyse it, and then give the concerns to the course organiser. But I don't think it's the job, of course, the organiser will add heavy things to him, and he will ignore it most of the time.

Participant 5:

This feedback will be implemented for the students with each course. I believe they will conduct it, I think, from next year. So, you give the feedback after the students finish the course which will include all of the assessments, the midterm and the final.

The researcher:

So, who is receiving the feedback, the students receiving feedback from you?

Participant 5:

No, no. The students are already giving feedback as a survey. They fill it out online. But, I believe next year, there will be feedback from the course organisers to the students after the midterm and after the final. I don't know; maybe Dr. (x) will confirm that.

Participant 3:

Yeah, It's not an individual assessment for the whole for every student. It's like feedback for the whole patch. Like the whole patch with the boys' patch, there would be feedback from the staff member of each unit or course or department.
I agree strongly with Dr. (x) regarding the feedback. It's only handled by the course director, which I think is really not appropriate. Because it has to go through Medical Education because they are the ones who are concerned regarding the assessment tools and method of teaching, because we do not have the expertise in this field as clinicians, and because we are not medical education experts, it must be integrated. This feedback is not really evaluated by medical education. I don't know what's the obstacle regarding this one. It's only, but just taken out and put in the course protocol or course things as just hard materials. Because of that, we are waiting for you to come.

The researcher:

Anything else?
Do you think that your OSCE has an educational impact?

Participant 2:

Yes. Actually, we work hard to make sure that this OSCE will improve the students when we are writing the OSCE, especially midterm OSCE because it will improve them for the rest of the course. So, we choose our questions and target questions regarding the more important subjects in their practice and the things that will make sure even if the student did not perform well in the exam. This exam will help the student and enable him to be sure that he will improve himself in this area.

Participant 4:

Yes. Yes. We certainly agree with this.

Participant 1:

For medical college students, I think if the OSCE is implemented in the middle of the course, OK, NOT at the end of the course, this will be an educational one for them. But at the end of the course, the student will be preoccupied with the grades or grade oriented, OK, but if it is implemented in the mid-term or during the course, it will be part of his education.

Participant 2:

Yes, in general surgery, yes, because I think I mean half of the training is hospital based. So, whatever they find deficiencies in them in the midterm, they try to correct it. It has an academic rule.

The researcher:

Nice.

Participant 3:

Yes, I agree as surprisingly, even in the deanship of academic affairs that during the sub-intern course this year, the surgical one for the boys, the grades were totally different from previous courses in the surgical department. Because they did the four subspecialty rotations that were not available in our college, their grades started to be different. For example, the one who got high marks in MCQs and OSCE got the worst or average grades during the sub-intern assessment. However, the one who will get bad when he gets higher. So, this means I think we have to look again into our assessment, So I agree better implementation for the OSCE in a more clinical way and more optimal way would be more helpful.

The researcher:

Very nice let's move to the next question, how do the college curriculum and the learning activities affect a student's performance on your OSCEs?
Is it clear?

Participant 5:

Could you elaborate on this more, please?

The researcher:

So how about the curriculum, the learning activities, the lecture, the bedside teaching? Is it oriented to your clinical assessment?

Participant 5:

Yes, actually, most of the scenarios in the OSCE would be already given to the students, either through lectures or bedside teaching.

Participant 4:

We teach the concepts, basic concepts, and from those concepts, OSCEs come. So, if they learn the concept OSCE becomes easy.

Participant 2:

So, as the course director of ENT and the chairman of the surgical department, I may disagree with orthopaedic, but I agree with general surgery. In orthopaedic, ENT, and ophthalmology, we are not doing most of the bedside teaching. It's very limited bedside teaching in orthopaedic than in ENT and ophthalmology. However, in general, surgery is still good. But suppose we are teaching our students in non-academic or in an unteaching hospital. In that case, there are some difficulties because there are some limitations or huge limitations, even bedside teaching. There is refusal, there are administration issues, or even there's, how can I say, a clinician interaction or clashes between them and our students and our staff. It's somewhat difficult because we have some difficulty with this one.
So, we are not doing the optimal teaching that allows us to give them the optimal OSCE, and there are still huge limitations in this one because I took most of this feedback from the students directly, which is what's really happening because even as a chairman, I'm talking with the vice dean of academic affairs and the vice dean of clinical affairs to solve this issue.

This issue It's happening in the whole (the place name) because it's not only in (the college name) College of Medicine, or in the main campus college in (the place name). They have the same issue with this one, bedside teaching.

Participant 6:

Sorry, I may understand that wrongly. Exactly the same thing happens as Dr. (x) and Dr. (x) said. I thought that the topics on the OSCE would already be explained to the students. That's what my understanding was. No, we face difficulties. It's not complete. I mean, the clinical practice will not be completely a mirror picture of what OSCE they've got NO. What we have in clinical and bedside teaching is limited because of the obstacles and limitations that are already mentioned.

The researcher:

Alright. So, do you provide written feedback or even oral feedback for students in each station? And why?

Participant 2:

No, we do not. I don't think we do.

Participant 5:

No, we don't. We don't do them.

Participant 6:

How can we do it?

Participant 5:

Because no one told us this is an important thing. That's the issue.

Participant 6:

The question is, how can you do it? Suppose, I mean, let me give you a practical example. I have a student. I examined him for an abdominal examination. I want to give him feedback after the station. Do you have a recording? I mean, we need a recorded version of it. How can we tell him why and where he went wrong? You cannot explain a clinical assessment in words. There should be a recording then, like in the West. They record the exam, and then they show the recorded exam to the student. This is the feedback, this is the mistake, and this needs correcting. How can we do it?

It's important, but how can you do it?

The researcher:

OK.

Participant 4:

I need to see them again. You need a recorded version to see so that we can show it. We cannot give him after the station because he has to go to another station. Suppose I give him bad feedback after a station. This would affect his performance at the next station. So, we had to do it after exam after exam. He must remember what he did; the only way is the recorded exam.

Participant 5:

So, Dr Khaled, do you mean general feedback or students' feedback?

The researcher:

Personal students' feedback?

Participant 5:

Oh, no. This is very difficult.

Participant 3:

So just let me. Let me just elaborate. I agree with Dr. (x) we're not doing this, although it's very important. I think it's my mind. Why aren't we doing it? Because first of all, it's not an ideal OSCE exam. Number two, there is no recording; thirdly, there's no structural feedback method for the students. And our students are not, I mean, trained regarding the OSCE regarding the feedback. We have never been taught how to get feedback. Normally they are super emotional, although there are many voices calling for that feedback.

The researcher:

Nice, I think. Is the same thing with other courses?

OK, I can move on to the next question. What do you consider to be the rule of the college and teachers in preparing the students for the OSCE?

Participant 4:

What I believe is that before the summative OSCE, there should be a demo OSCE. We should show students a sample of it. So, they did not get shocked by the exam, this was the first time they were bearing an exam with marks, which had grades that could fail them or pass, and it was the first exposure before the exposure. There should be a rehearsal, so there should be a sample of it which is ungraded.

Participant 3:

Dr Khaled.

The researcher:

Yes, please.

Participant 3:

Yes, Dr. (x) continue, sorry!

Participant 6:

I'm done, I'm done. This is all I have to say.

Participant 3:

OK, Dr. Khaled, you asked us about what's the role of college or our role in college.

The researcher:

The role of the college is to provide, let's say, the regulations, the financial support, and the support in general, as well as your role in helping the students do good in OSCE.

Participant 3:

OK, so if we talk about the whole college, there is good support regarding the external examiners; they will provide whatever is needed. Regarding the fund, there is a limitation on the funds. This is from the university administration, not from the college itself. However, the college is like giving an affiliated academic title to those who help in writing the OSCEs or as external examiners. So, regarding general surgery, we are doing our best, but we still need more.

Participant 2:

We need to encourage them to come to the college and assess. I agree with Dr. (x) when he said there is no training for the students at all. I think we are supposed to give them more training on the clinical exams before they embark on them, or they implement it in their basic science.

The researcher:

You have mentioned many opportunities, many challenges, actually. So how would you conduct OSCE if you had more resources and authority?

Participant 4:

We will do it in an ideal way.

The researcher:

So, can you explain the ideal way briefly?

Participant 4:

I mean, suppose in general surgery, we have a dream that we should have a very big hall examination hall. We have at least 12 stations. We should have a timekeeper. The students, 12 students will enter. Then every five minutes, they would rotate, then when the bell rang, the student moved. I mean, just the students will move and have a paper in their hands. And in the station, they fill in their response. Nobody is talking to anybody. Then five or six stations, there will be objects, they'll be specimens. In another five or six stations, there'll be skills; like in real life, it is doable. It is not something that is not possible. We have been through it. Inshallah.

Inshallah.

Participant 3:

I agree with Dr. (x) we are trying our best to get the college to implement this once we move to the new building; maybe we can optimise the building to conduct the OSCE exams.

Participant 5:

But Dr Khaled, I will add to that good clinical training and teaching for the students before the exam. I mean, we need better situations in the hospital, teaching hospital, bigger wards, different varieties of patients, let's say in orthopaedics. All the patients are trauma patients. We cannot explain or discuss with the students' sports-related or trauma-related pathologies. It's only trauma. So, a variety of patients must be there. We need to have real clinical patients, not only simulations. So, these may be improving in the future.

The researcher:

Anything Dr. (x) or Dr. (x) You want to add?

Participant 2:

No, nothing to add. I just want to emphasise that a proper place for the OSCE is very, very important. And we lack the proper place. But still, we have a good place for conducting OSCE. We hope that with the new building, this will be better so we can conduct it better. Because, you know, we have renovated the old building to conduct the OSCE and still it's not usually used. So usually, we have to face logistic issues such as cleaning the place and providing good service during the OSCE exam.

Also, staff dedication is very important, and we are lacking that. And good training for the staff and simulated patients as well.

The researcher:

Alright. So, it comes to my mind now, as it is a newly established medical school. Is that adding an extra layer of challenges to you? And so, what's your opinion about that? Is it a good or bad thing to be a newly established medical school?

Participant 4:

It's good.

Participant 1:

It's not about starting a new college, but I think it's related to the funds that you have. If you have the logistics, as my colleagues said before, and the funds to conduct a really good OSCE, anyone can do it then, a new or an old college. But you have to have the facilities and the funds for that. As long as we do not have a university hospital and a limited number of patients, we need specialised models and mannequins to run the exam, which are also not always available. It costs a lot of money. Also, the university hospital will provide a real environment and enough rooms to conduct enough OSCE stations.

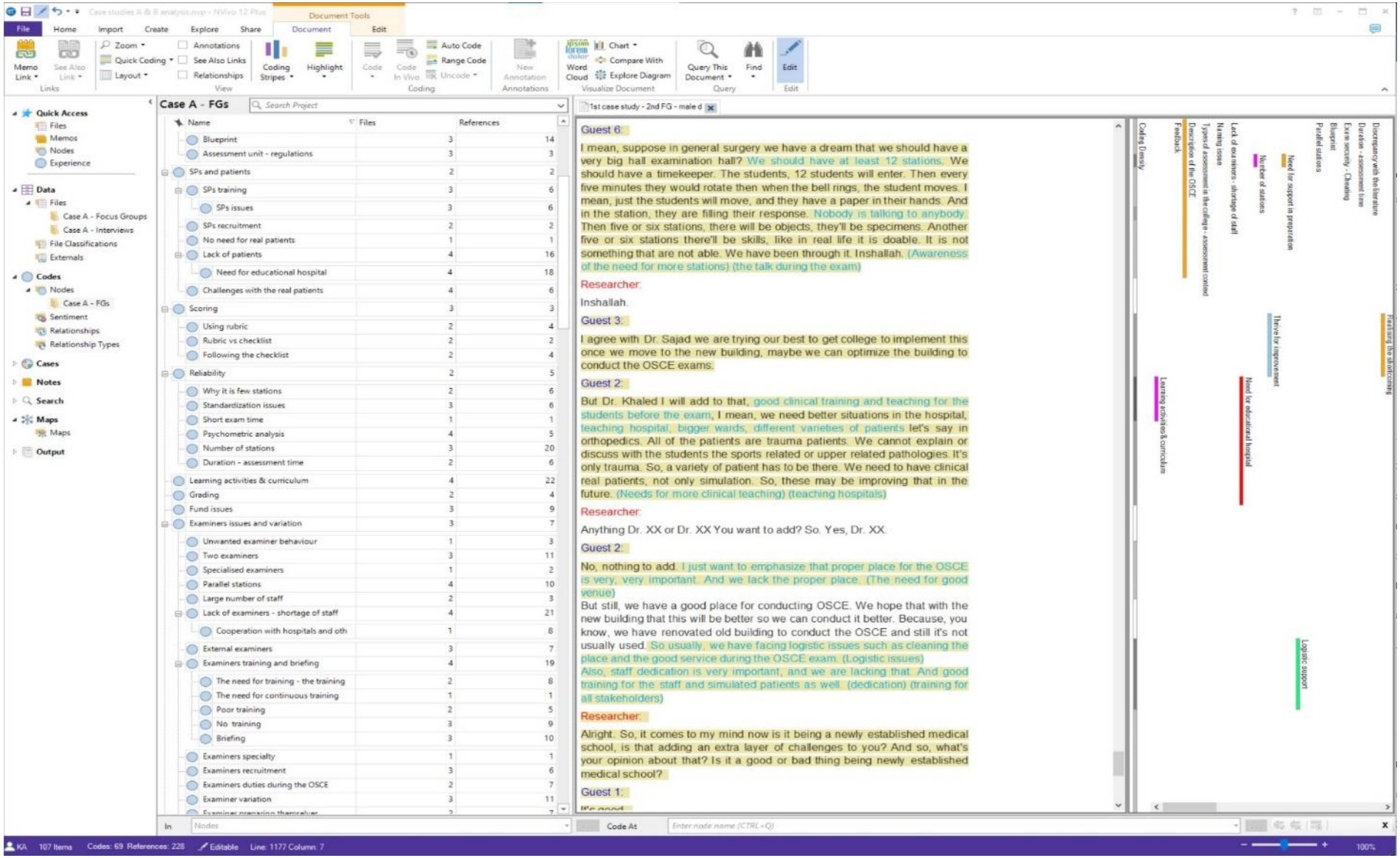
Participant 6:

I think the issue has many potential reasons. I believe being newly established is just one of them. For example, the regulations are not mature enough, and the infrastructure is not adequate to do what we want to do.

Participant 2:

I agree it's not related to being a new college because we have many new colleges in Saudi Arabia, and they are doing great regarding the OSCE and the exam. Our issue here is the funds because we have two colleges in the same university.

Appendix 26. Screenshots depict the codes generation process using NVivo.

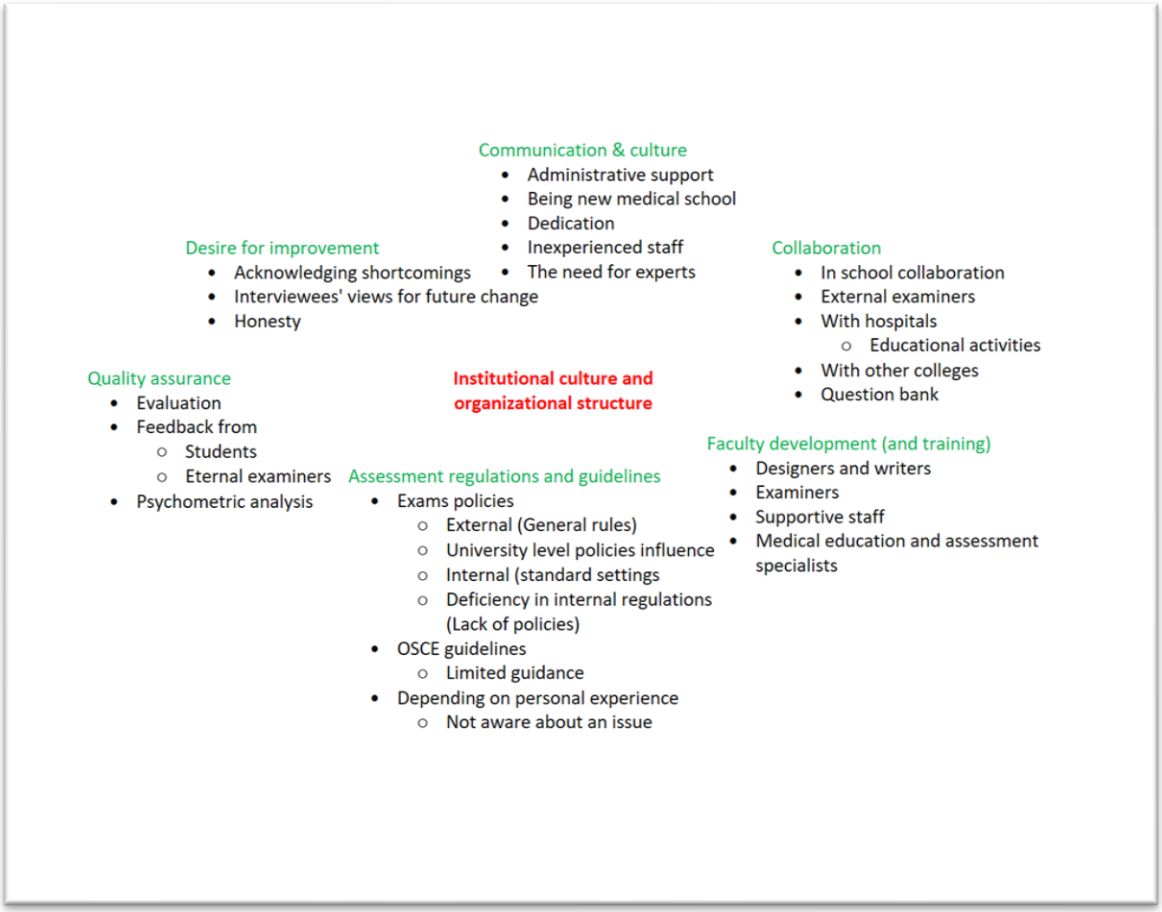


Appendix 27. The initial codes from the four interviews and four FGs of case study A.

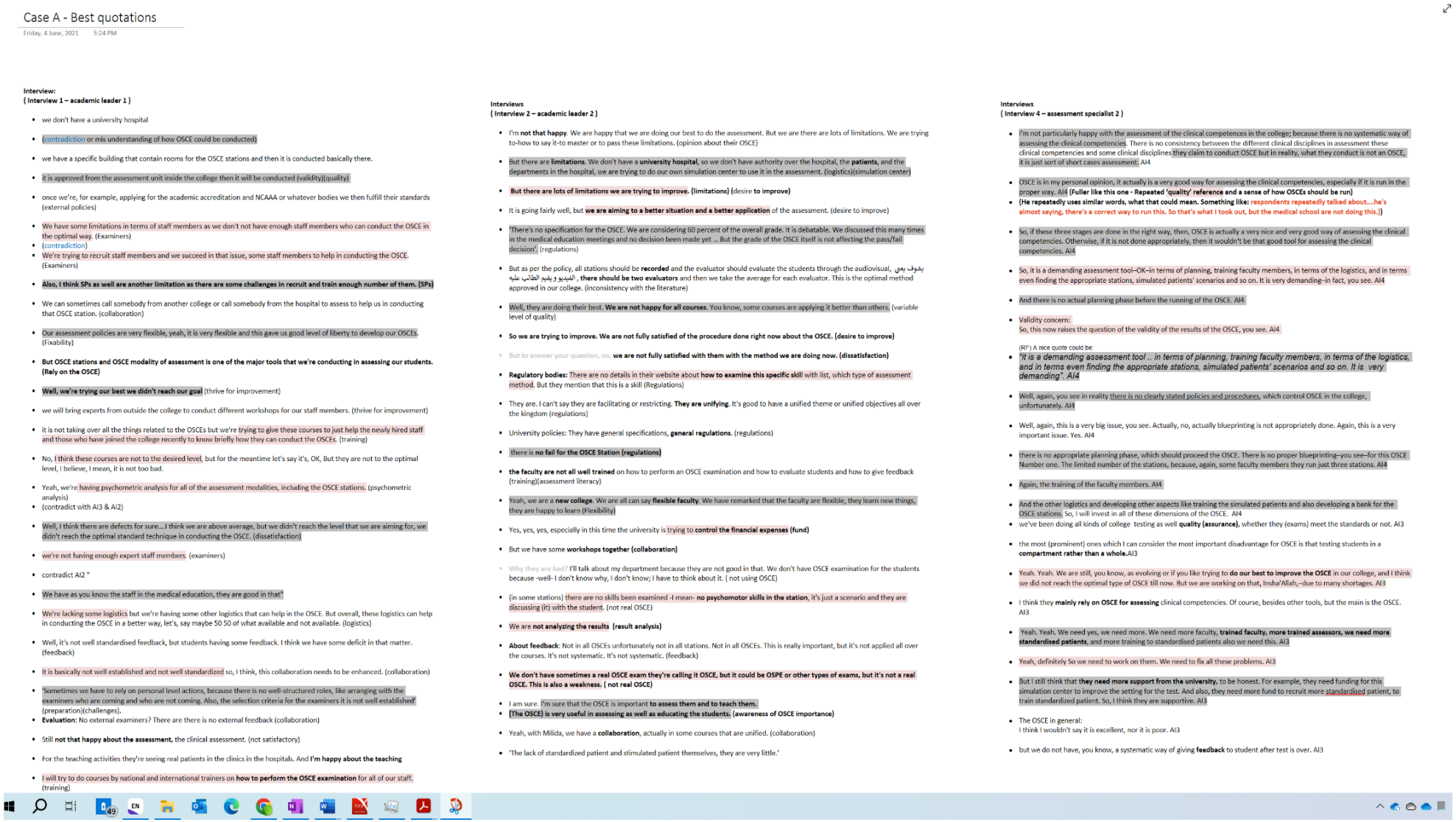
The initial codes	Evidence/References
Background	6
Communication - in schools' collaboration	13
Examiner and designer collaboration	3
Communication & culture	6
Administrative support	1
Being a new medical school	3
Dedication	5
Inexperienced	7
The need for collaboration	6
Lack of resources implications	3
Naming issue	18
Need for support in preparation	17
Opportunities	3
Realising the shortcoming	7
Staff opinion on the OSCE - as a good tool	13
Inauthentic - Authentic	6
Knowledgeable	3
Thrive for improvement	9
The language is a barrier	3
Types of assessment in the college - assessment context	16
Regulations & policies	5
External regulations/Rules	13
benefits	2
freedom	6
General rules	3
Limited guidance	7
National OSCE	2
Internal regulations	8
Cut-off score	13
Deciding the type of assessment	3
Deficiency in internal regulations - missing policies	8
Freedom & Flexibility	10
University-level policies influence	7
Description of the OSCE	13
Depending on personal experience	18
Discrepancy with the literature	11
Exam committee	3
Exam security - Cheating	8
Hospital complexity	2
Logistic support	12
Organisers dissatisfaction	5
Parallel stations - time and reliability	8
Preparation and planning	12
Planning	3
Supporting staff	2
The need for good design	4
Venue - Simulation centre	17
Virtual OSCE	1
Design choices	5
Disregard - Ignorance	7
Experience in designing the OSCE	6
Guideline - Lack of guidelines	9
Regulations	4
Models and mannequins available	3
OSCE station types	5
Spot diagnosis	2
Station writers	12
Educational impact	15
Feedback	15
Formative OSCE	6

Student's training	7
In future practice	9
Examiners issues and variation	7
Checklist objectivity vs subjectivity - Negligence	8
Checklist development	15
Examiner's exhaustion - tiredness	6
Examiner interaction with students	16
Examiners preparing themselves	7
Examiner's variation	15
Examiner's duties during the OSCE	7
Examiner's recruitment	8
Examiner's speciality	2
Examiner's training and briefing	19
Briefing	10
No training	20
Poor training	5
The need for continuous training	4
The need for training - the training received	12
External examiners	7
Lack of examiners - shortage of staff	21
Cooperation with hospitals and other schools	8
Large number of staff	3
Parallel stations	10
Specialised examiners	4
Two examiners	11
Unwanted examiner behaviour	3
Fund issues	9
Grading - standard setting	7
Learning activities & curriculum	22
Reliability	5
Duration - assessment time	6
Number of stations	20
Psychometric analysis	5
Short exam time	4
Standardisation issues	8
Why it is few stations	9
Scoring	3
Following the checklist	7
Rubric vs checklist	4
Using rubric	5
SPs and patients	5
Challenges with the real patients	6
Lack of patients	16
Need for educational hospital	18
No need for real patients	5
SPs recruitment	6
SPs training	6
SPs issues	7
Validity issues	6
Assessment unit - regulations	3
Blueprint	14
Dependent stations - validity issue	1
Exam committee	4
Exam supervisor	2
External support	7
Fairness	11
Organiser’s training	6
Quality assurance - Evaluation	17
Stations writers	14
The need for more preparation	7

Appendix 28. One of the earliest cycles for developing themes and subthemes. Themes in red, subthemes in green, and codes in black.



Appendix 29. A sample of my strategy for collecting the best quotations during the RTA analysis.



Case A - Best quotations

Friday, 4 June, 2021 5:24 PM

Interview:
(Interview 1 – academic leader 1)

- we don't have a university hospital
- (contradiction or mis understanding of how OSCE could be conducted)
- we have a specific building that contain rooms for the OSCE stations and then it is conducted basically there.
- it is approved from the assessment unit inside the college then it will be conducted (validity)(quality)
- once we're, for example, applying for the academic accreditation and NCAAA or whatever bodies we then fulfill their standards (external policies)
- We have some limitations in terms of staff members as we don't not have enough staff members who can conduct the OSCE in the optimal way. (Examiners)
- (contradiction)
- We're trying to recruit staff members and we succeed in that issue, some staff members to help in conducting the OSCE. (Examiners)
- **Also, I think SPs as well are another limitation as there are some challenges in recruit and train enough number of them. (SPs)**
- We can sometimes call somebody from another college or call somebody from the hospital to assess to help us in conducting that OSCE station. (collaboration)
- Our assessment policies are very flexible, yeah, it is very flexible and this gave us good level of liberty to develop our OSCEs. (Fixability)
- **But OSCE stations and OSCE modality of assessment is one of the major tools that we're conducting in assessing our students. (Rely on the OSCE)**
- **Well, we're trying our best we didn't reach our goal** (thrive for improvement)
- we will bring experts from outside the college to conduct different workshops for our staff members. (thrive for improvement)
- it is not taking over all the things related to the OSCEs but we're trying to give these courses to just help the newly hired staff and those who have joined the college recently to know briefly how they can conduct the OSCEs. (training)
- No, I think these courses are not to the desired level, but for the meantime let's say it's, OK, But they are not to the optimal level, I believe, I mean, it is not too bad.
- Yeah, we're having psychometric analysis for all of the assessment modalities, including the OSCE stations. (psychometric analysis)
- (contradict with AI3 & AI2)
- Well, I think there are defects for sure....I think we are above average, but we didn't reach the level that we are aiming for, we didn't reach the optimal standard technique in conducting the OSCE. (dissatisfaction)
- we're not having enough expert staff members. (examiners)
- contradict AI2 "
- We have as you know the staff in the medical education, they are good in that"
- We're lacking some logistics but we're having some other logistics that can help in the OSCE. But overall, these logistics can help in conducting the OSCE in a better way, let's, say maybe 50 50 of what available and not available. (logistics)
- Well, it's not well standardised feedback, but students having some feedback. I think we have some deficit in that matter. (feedback)
- It is basically not well established and not well standardized so, I think, this collaboration needs to be enhanced. (collaboration)
- 'Sometimes we have to rely on personal level actions, because there is no well-structured roles, like arranging with the examiners who are coming and who are not coming. Also, the selection criteria for the examiners it is not well established' (preparation)(challenges).
- **Evaluation**: No external examiners? There are there is no external feedback (collaboration)
- Still **not that happy about the assessment,** the clinical assessment. (not satisfactory)
- For the teaching activities they're seeing real patients in the clinics in the hospitals. And **I'm happy about the teaching**
- I will try to do courses by national and international trainers on **how to perform the OSCE examination** for all of our staff. (training)

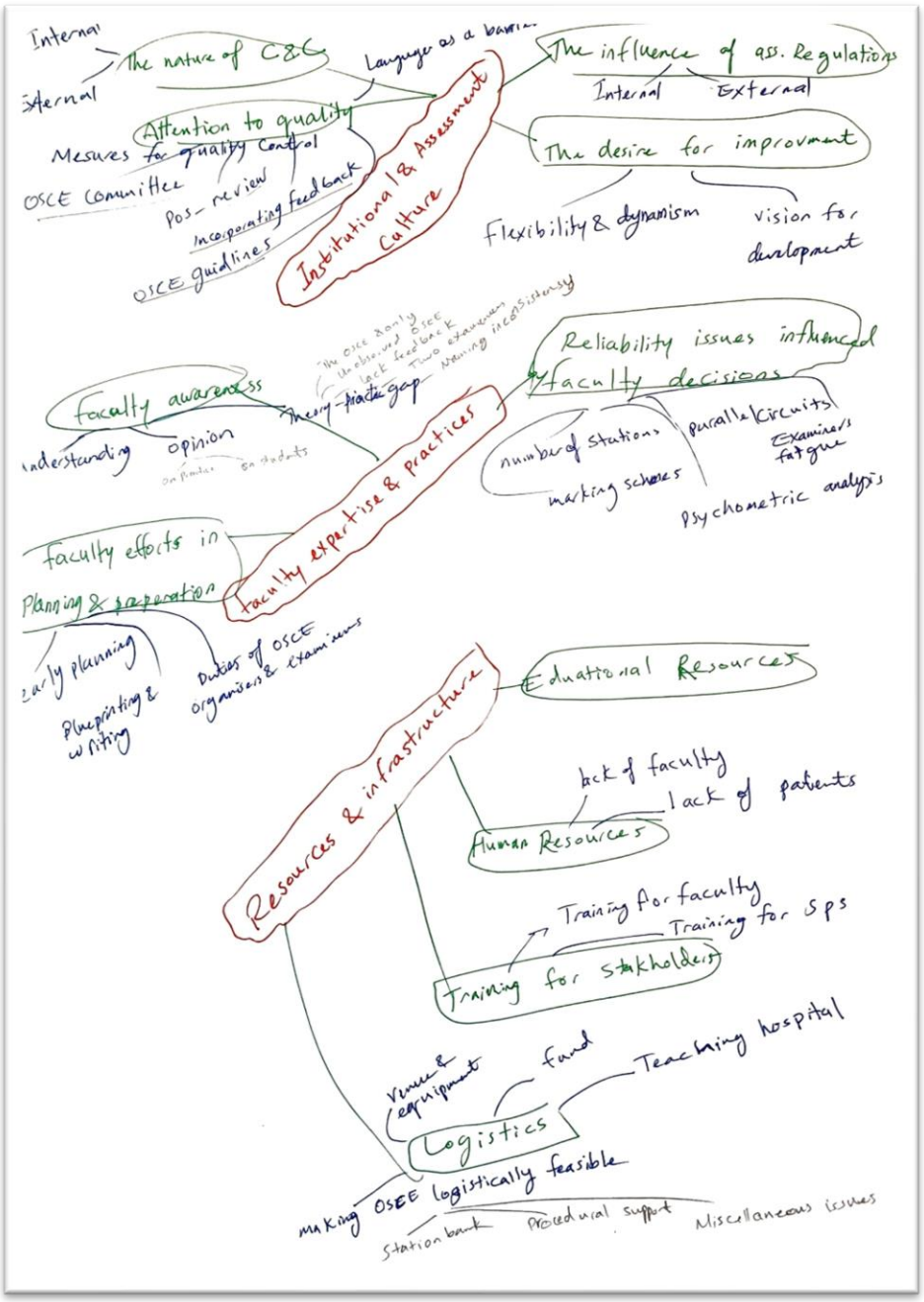
Interviews
(Interview 2 – academic leader 2)

- I'm **not that happy**. We are happy that we are doing our best to do the assessment. But we are there are lots of limitations. We are trying to-how to say it-to master or to pass these limitations. (opinion about their OSCE)
- But there are **limitations**. We don't have a **university hospital**, so we don't have authority over the hospital, the **patients**, and the departments in the hospital, we are trying to do our own simulation center to use it in the assessment. (logistics)(simulation center)
- **But there are lots of limitations we are trying to improve. (limitations) (**desire **to improve)**
- It is going fairly well, but **we are aiming to a better situation and a better application** of the assessment. (desire to improve)
- 'There's no specification for the OSCE. We are considering 60 percent of the overall grade. It is debatable. We discussed this many times in the medical education meetings and no decision been made yet ... But the grade of the OSCE itself is not affecting the pass/fail decision'. (regulations)
- But as per the policy, all stations should be **recorded** and the evaluator should evaluate the students through the audiovisual, يخوف يعني الفيديو و يقيم الطالب عليه , **there should be two evaluators** and then we take the average for each evaluator. This is the optimal method approved in our college. (inconsistency with the literature)
- Well, they are doing their best. **We are not happy for all courses**. You know, some courses are applying it better than others. (variable level of quality)
- **So we are trying to improve. We are not fully satisfied of the procedure done right now about the OSCE. (desire to improve)**
- But to answer your question, no, **we are not fully satisfied with them with the method we are doing now. (dissatisfaction)**
- **Regulatory bodies:** There are no details in their website about **how to examine this specific skill** with list, which type of assessment method. But they mention that this is a skill (Regulations)
- They are. I can't say they are facilitating or restricting. **They are unifying**. It's good to have a unified theme or unified objectives all over the kingdom (regulations)
- University policies: They have general specifications, **general regulations**. (regulations)
- there is **no fail for the OSCE Station (regulations)**
- **the faculty are not all well trained** on how to perform an OSCE examination and how to evaluate students and how to give feedback (training)(assessment literacy)
- Yeah, we are a **new college**. We are all can say **flexible faculty**. We have remarked that the faculty are flexible, they learn new things, they are happy to learn (Flexibility)
- Yes, yes, yes, especially in this time the university is trying to **control the financial expenses (fund)**
- But we have some **workshops together (collaboration)**
- Why they are bad? **I'll talk about my department because they are not good in that. We don't have OSCE examination for the students** because -well- I don't know why, I don't know; I have to think about it. (not using OSCE)
- (in some stations) there are no skills been examined -I mean- **no psychomotor skills in the station**, it's just a scenario and they are discussing (it) with the student. (not real OSCE)
- We are **not analyzing the results (result analysis)**
- **About feedback**: Not in all OSCEs unfortunately not in all stations. Not in all OSCEs. This is really important, but it's not applied all over the courses. It's not systematic. It's not systematic. (feedback)
- **We don't have sometimes a real OSCE exam they're calling it OSCE, but it could be OSPE or other types of exams, but it's not a real OSCE. This is also a weakness. (not real OSCE)**
- I am sure. I'm sure that the OSCE is important **to assess them and to teach them.**
- **(The OSCE) is very useful in assessing as well as educating the students. (awareness of OSCE importance)**
- Yeah, with Milida, we have a **collaboration**, actually in some courses that are unified. (collaboration)
- 'The lack of standardized patient and stimulated patient themselves, they are very little.'

Interviews
(Interview 4 – assessment specialist 2)

- 'I'm not particularly happy with the assessment of the clinical competences in the college; because there is no systematic way of assessing the clinical competencies. There is no consistency between the different clinical disciplines in assessment these clinical competencies and some clinical disciplines they claim to conduct OSCE but in reality, what they conduct is not an OSCE, it is just sort of short cases assessment. AI4
- OSCE is in my personal opinion, it actually is a very good way for assessing the clinical competencies, especially if it is run in the proper way. AI4 **(Fuller like this one - Repeated 'quality' reference and a sense of how OSCEs should be run)**
- **(He repeatedly uses similar words, what that could mean. Something like: respondents repeatedly talked about....he's almost saying, there's a correct way to run this. So that's what I took out, but the medical school are not doing this.))**
- So, if these three stages are done in the right way, then, OSCE is actually a very nice and very good way of assessing the clinical competencies. Otherwise, if it is not done appropriately, then it wouldn't be that good tool for assessing the clinical competencies. AI4
- So, it is a demanding assessment tool–OK–in terms of planning, training faculty members, in terms of the logistics, and in terms even finding the appropriate stations, simulated patients' scenarios and so on. It is very demanding–in fact, you see. AI4
- And there is no actual planning phase before the running of the OSCE. AI4
- Validity concern:
  So, this now raises the question of the validity of the results of the OSCE, you see. AI4
- (RF) A nice quote could be:
- *"it is a demanding assessment tool .. in terms of planning, training faculty members, in terms of the logistics, and in terms even finding the appropriate stations, simulated patients' scenarios and so on. It is very demanding". AI4*
- Well, again, you see in reality there is no clearly stated policies and procedures, which control OSCE in the college, unfortunately. AI4
- Well, again, this is a very big issue, you see. Actually, no, actually blueprinting is not appropriately done. Again, this is a very important issue. Yes. AI4
- there is no appropriate planning phase, which should proceed the OSCE. There is no proper blueprinting–you see–for this OSCE Number one. The limited number of the stations, because, again, some faculty members they run just three stations. AI4
- Again, the training of the faculty members. AI4
- And the other logistics and developing other aspects like training the simulated patients and also developing a bank for the OSCE stations. So, I will invest in all of these dimensions of the OSCE. AI4
- we've been doing all kinds of college testing as well **quality (assurance),** whether they (exams) meet the standards or not. AI3
- the most (prominent) ones which I can consider the most important disadvantage for OSCE is that testing students in a **compartment rather than a whole.**AI3
- Yeah. Yeah. We are still, you know, as evolving or if you like trying to **do our best to improve the OSCE** in our college, and I think we did not reach the optimal type of OSCE till now. But we are working on that, Insha'Allah,–due to many shortages. AI3
- I think they **mainly rely on OSCE for assessing** clinical competencies. Of course, besides other tools, but the main is the OSCE. AI3
- Yeah. Yeah. We need yes, we need more. We need more faculty, **trained faculty, more trained assessors, we need more standardised patients**, and more training to standardised patients also we need this. AI3
- Yeah, definitely So we need to work on them. We need to fix all these problems. AI3
- But I still think that **they need more support from the university,** to be honest. For example, they need funding for this simulation center to improve the setting for the test. And also, they need more fund to recruit more standardised patient, to train standardized patient. So, I think they are supportive. AI3
- The OSCE in general:
  I think I wouldn't say it is excellent, nor it is poor. AI3
- but we do not have, you know, a systematic way of giving **feedback** to student after test is over. AI3

Appendix 30. One of the later cycles for theme and subtheme development.



Appendix 31. The abstract for the oral presentation presented at AMEE (2020).

Essential Components for Implementing High-Quality OSCEs: A Scoping Review

Khaled Almisnid[a,c], Matt Homer[a], Richard Fuller[b], and Trudie Roberts[a]

[a] Leeds Institute of Medical Education, School of Medicine, University of Leeds, Leeds, UK.
[b] School of Medicine, University of Liverpool, Liverpool, UK.
[c] Medical Education Unit, Unaizah College of Medicine, Qassim University, KSA.

Type of presentation: Poster

ABSTRACT

Background

High-quality clinical assessments are a crucial component of any medical school's assessment programme. The Objective Structured Clinical Examination (OSCE) is one of the most common instruments used worldwide to assess medical students' clinical skills and competencies. The OSCE is widely adopted for high-stakes exams, yet it is often implemented variably, with few measures of assessment 'quality', across institutions. Identifying the OSCE components that consistently produce high-quality clinical assessments would allow the development and implementation of valid and reliable OSCEs.

Summary of work

A scoping review was conducted to explore the available literature concerning OSCE design and implementation. The search terms were: *OSCE, (quality, implementation, design or application), medicine* and *undergraduate*. For inclusion, articles were required to be written in English, have search terms in the titles or abstracts, and discuss the OSCE implementation and design process. The 'advanced search' option was used to combine search terms in nine databases: PubMed, Ovid Embase, Global Health, PsycINFO, MEDLINE, Web of Science, Cochrane Library, ERIC and Scopus. The reference list of each included article was manually searched for additional articles that also met the scoping review criteria (see appendix).

Results and Discussion

The search found 554 publications, but only 39 met the inclusion criteria. Analysis of these papers revealed that the common key components essential for producing high-quality OSCEs included: establishing an examination planning board; conducting thorough blueprinting; training the exam organisers, station writers, and assessors; training simulated/standardised patients; using validated scoring sheets; planning the venue logistics; undertaking post hoc psychometric analysis; evaluating the implementation process; and giving and receiving feedback from different stakeholders to inform development of subsequent implementations.

Conclusion

Implementing high-quality OSCEs is a complex task – a multifactorial process in which each component can affect the overall quality of the assessment. The results of this scoping review suggest a set of common components that, when appropriately employed, could help designers implement high-quality OSCEs.

Take home message

Each of the components mentioned above should be considered and employed appropriately and professionally during the implementation process to ensure high-quality OSCE.

Appendix 32. A sample from the workshop presented at a medical school in SA (2021).

The OSCE
in Theory

Khaled Almisnid

Principles of good assessments

Validity

The assessment measures what it is supposed to measure and reflects the curriculum.

Reliability

When the same assessment is introduced again (in similar circumstances), it should produce similar results.

Feasibility

The practicality of assessment in terms of resources required for implementation.

Principles of good assessments

Educational and Catalytic Impact

The educational benefit that students gain from taking assessment.
The motivation that all stakeholders gain from assessment results.

Fairness

The assessment results lead to similar decisions across equivalent cohorts or institutions.

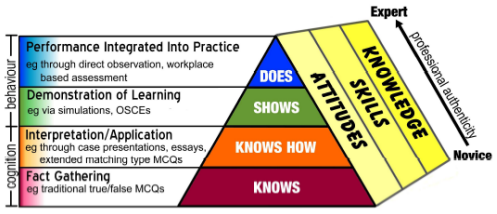
Acceptability

Different stakeholders recognise assessment importance and credibility.

Why OSCE is Important

- In any assessment programme multiple tools should be used.
  - System of assessment.
  - OSCE is only one unit.
- Assessment drives learning VS Learning drives assessment.

Miller's Pyramid



Based on work by Miller GE, The Assessment of Clinical Skills/Competence/Performance; Acad. Med. 1990; 65(9); 63-67
Adapted by Drs. R. Mehay & R. Burns, UK (Jan 2009)

Keep in mind

Designing the OSCE is an iterative rather than a linear process.

Implementing high-quality OSCEs is a complex task – a multifactorial process in which each component can affect the overall quality of the assessment.

Therefore, each of these components should be considered and employed appropriately and professionally during the implementation process to ensure a high-quality OSCE.

Appendix 33. The poster for the Saudi International Medical Education Conference (SIMEC) (2022).

SIMEC 2022
The Saudi International Medical Education Conference
المؤتمر السعودي الدولي للتعليم الطبي

The Necessary Elements for Developing Successful Objective Structured Clinical Examinations

Khaled Almisnid[1,3*], Matt Homer[1], Trudie Roberts[1], and Richard Fuller[2]

[1] Leeds Institute of Medical Education, School of Medicine, University of Leeds, Leeds, UK.
[2] School of Medicine, University of Liverpool, Liverpool, UK.
[3] Medical Education Unit, Unaizah College of Medicine, Qassim University, KSA.

UNIVERSITY OF LEEDS

Qassim University جامعة القصيم

Abstract Id no SIMEC2022-126

INTRODUCTION

High-quality clinical assessments are a crucial component of any medical school's assessment programme. The Objective Structured Clinical Examination (OSCE) is one of the most common instruments used worldwide to assess medical students' clinical skills and competencies. The OSCE is widely adopted for high-stakes exams, yet it is often implemented variably, with few measures of assessment 'quality', across institutions. Identifying the OSCE components that consistently produce high-quality clinical assessments would allow the development and implementation of valid and reliable OSCEs.

AIM

Identifying the OSCE components that consistently produce high-quality clinical assessments would allow the development and implementation of valid and reliable OSCEs.

METHODOLOGY

- A scoping review was conducted to explore the available literature concerning OSCE design and implementation.
- The search terms were: OSCE, (quality, implementation, design or application), medicine and undergraduate.
- For inclusion, articles were required to be written in English, have search terms in the titles or abstracts, and discuss the OSCE implementation and design process.
- The 'advanced search' option was used to combine search terms in nine databases: PubMed, Ovid Embase, Global Health, PsycINFO, MEDLINE, Web of Science, Cochrane Library, ERIC and Scopus.
- The reference list of each included article was manually searched for additional articles that also met the scoping review criteria.

The included papers:

- Papers that contains opinions and experience of well known OSCE experts.
- Papers that demonstrate high-quality OSCEs by:
  - Showing valid OSCEs by detailing their good practise in OSCE implementation.
  - Showing reliable OSCEs results by conducting psychometric analysis.

RESULTS

- The search found 554 publications, but only 39 met the key inclusion criteria.
- Analysis of these papers revealed that the common key components essential for producing high-quality OSCEs included:

RESULTS

1. establishing an examination planning board;
2. conducting thorough blueprinting;
3. training the exam organisers, station writers, and assessors;
4. training simulated and standardised patients;
5. using validated scoring sheets;
6. planning the venue logistics;
7. undertaking post hoc psychometric analysis;
8. evaluating the implementation process;
9. giving and receiving feedback from different stakeholders to inform development of subsequent implementations.

CONCLUSION

Implementing high-quality OSCEs is a complex task – a multifactorial process in which each component can affect the overall quality of the assessment. The results of this scoping review suggest a set of common components that, when appropriately employed, could help designers implement high-quality OSCEs.

REFERENCES

- Boursicot, K. and Roberts, T. 2005. How to set up an OSCE. *The clinical teacher.* 2(1), pp.16-20.
- Casey, P.M., Goepfert, A.R., Espey, E.L., Hammoud, M.M., Kaczmarczyk, J.M., Katz, N.T., Neutens, J.J., Nuthalapaty, F.S., Peskin, E., Assoc Prof, G. and Obstet, U. 2009. To the point: reviews in medical education-the Objective Structured Clinical Examination. *American Journal of Obstetrics and Gynecology.* 200(1), pp.25-34.
- Daniels, V.J. and Pugh, D. 2018. Twelve tips for developing an OSCE that measures what you want. *Medical teacher.* 40(12), pp.1208-1213.
- Fuller, R., Homer, M., Pell, G. and Hallam, J. 2017. Managing extremes of assessor judgment within the OSCE. *Medical teacher.* 39(1), pp.58-66.
- Harden, Lilley, P. and Patricio, M. 2016. *The Definitive Guide to the OSCE: The Objective Structured Clinical Examination as a performance assessment.* 1 ed. Elsevier Health Sciences.
- Harden, R.M. 1990. Twelve tips for organizing an Objective Structured Clinical Examination (OSCE). *Medical Teacher.* 12(3-4), pp.259-264.
- Hodges, B. 2003. OSCE! Variations on a theme by Harden. *Medical Education.* 37(12), pp.1134-1140.
- Khan, K.Z., Ramachandran, S., Gaunt, K. and Pushkar, P. 2013. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: An historical and theoretical perspective. *Medical Teacher.* 35(9), pp.e1437-e1446.
- Pell, G., Fuller, R., Homer, M. and Roberts, T. 2010. How to measure the quality of the OSCE: A review of metrics AMEE guide no. 49. *Medical Teacher.* 32(10), pp.802-811.

ACKNOWLEDGMENT

I would thank Dr Matt Homer, professor Trudie Robert, and Professor Richard Fuller for their continuous support thorough all stages of this research.

Appendix 34. This table summarises the themes, subthemes, and subheadings of case study A with example quotes.

Themes	Subthemes	Subheadings	Example quote
First theme **Institutional and assessment culture**	The nature of communication and collaboration	Internal communication	*'Other parties like staff and departments are not willing to communicate or help experiment with something new...' (FG1-D2).*
		External collaboration	*'It is basically not well established and not well standardised, so I think this collaboration needs to be enhanced.' (Interview1-AL).*
	The Influence of assessment regulations	External and internal regulations	*'For the OSCE, we do not have any guideline. However, if it is there, we are not aware of it.' (FG1-D2).*
		Dependency on individual experience	*'Honestly, yes, it is all my experience.' (FG3-E5).*
	The desire for improvement	Being a new medical school	*'But to run a new medical college without having clearly stated policies and procedures…This is a problem.' (Interview4-AS).*
		Vision for development	*'If I had more authority, then I would invest in running an appropriate OSCE in our college. Because, again, all that really matters is certifying competent doctors.' (Interview4-AS).*
Second theme **Faculty expertise and practices**	Faculty awareness of the OSCE	Faculty understanding	*'The OSCE reflects the students' clinical skills, but from another point, it does not reflect the reality.' (FG3-E5).*
		Faculty opinion on their OSCE	*'I am not very happy; however, I think we are doing the best OSCE that we can produce based on the resources we have.' (FG4-E3).*
	Faculty approaches in planning and preparation	Preparation and writing of the stations	*'There is no appropriate planning phase that should precede the OSCE. For instance, there is no proper blueprinting.' (Interview4–AS).*
		Examiners suggestions for improvement to the OSCE	*'The examiner needs to know her station one or two days ahead, to prepare herself and discuss any ambiguity. This could enable us to perform better.' (FG3-E3).*
	Quality issues	Challenges to quality assurance	*'The OSCE Committee get all points of view, then the good feedback incorporated for improvement.' (FG4-E5).*
		The language as a barrier	*'English as a language for communication in the OSCE stations, is limiting students' ability to express their thoughts and knowledge properly.' (FG1-D2).*

Third theme **The OSCE quality and design**	Theory-practice gap	Only the OSCE	*'I think they mainly rely on the OSCE for assessing clinical competencies… Other tools are formative'. (Interview3-AS).*
		OSCE design confusion	*'We do have a written OSCE for the midterm, ok. It is a written OSCE case with questions about the diagnosis and management.' (FG4-E2).*
		Two examiners	*'In each station, we have two examiners, one internal from the college and one external from the hospital.' (FG1-D3).*
		Unwanted examiner behaviours	*'The examiners keep asking the student during the station about questions or skills that are not mentioned in the checklist.' (FG3-E3).*
		Lack of individual feedback	*'We still do not give feedback to the students.' (Interview3-AS).*
	Reliability issues	Number of stations	*'Three stations. Can a three-station OSCE be considered an OSCE?' (Interview4-AS).*
		Parallel circuits and examiners' challenges	*'We may need two tracks in all exams, but we do not have enough examiners.' (FG3-E6).*
		Marking schemes design	*'If we could have a well-built rubric or checklist, it would minimise the bias or the variation between the benign and malignant assessors, so to speak.' (FG1-D1).*
Fourth theme **Resources and infrastructure setup**	Human resources	Lack of workforce	*'We need instructors and experts who have a good background in how to do things (in assessment).' (FG1-D1).*
		Training for stakeholders	*'There is no training here for the faculty about how to introduce or conduct an OSCE exam in the proper way… Actually, there is no training.' (FG2-D1).*
	Instructional and logistical resources	Educational resources	*'I would say that this is a real strength in our curriculum. It has an effect on how we use OSCE in our college.' (Interview3-AS).*
		Logistical resources	*'I just want to emphasise that the proper place for the OSCE is very, very important. And we lack the proper place.' (FG2-D4).*

Appendix 35. This table summarises the themes, subthemes, and subheadings of case study B with example quotes.

Themes	Subthemes	Subheadings	Example quote
First theme **Institutional and assessment culture**	The nature of communication and collaboration	Internal communication	'This is teamwork, actually; no one of us can do it all. We collaborate from writing the scenario all the way to arranging all material…' *(FG4-E1).*
		External collaboration	'We cooperate with outside examiners who come in.' *(Interview3-AS).*
	The Influence of assessment regulations	External regulations	'We responded very well to the NCAAA recommendations.' *(Interview2-AS).*
		Internal regulations	'Our policies are all clear and updated.' *(Interview1-AL).*
	The desire for improvement	Flexibility and adaptability	'We are receptive to learning and improving things. This is life. No one is perfect.' *(FG1-D4).*
		Vision for development	*'We also need to think bigger by merging the Assessment Unit with the OSCE Committee to establish an Assessment and Examination Centre.'* (Interview2-AS).
Second theme **Faculty expertise and practices**	Faculty awareness of the OSCE	Faculty understanding	*'The OSCE needs a good design to be effective.' (Interview1-AL).*
		Faculty opinion on their OSCE	*'I think it has sort of good quality but not of the top quality.' (FG3-E2).*
	Faculty efforts in planning and preparation	Early planning and blueprinting	*'One of the strengths is good early planning. We have at least two to three months of preparation.' (FG4-E5).*
		Duties of the OSCE organisers and examiners	*'Before the exam itself, we have a calibration meeting for the examiners.' (Interview3-AS).*
	Attention to quality	Measures for quality control	*'The OSCE Committee get all points of view, then the good feedback incorporated for improvement.' (FG4-E5).*
		The language as a barrier	*'Simulated patient and student communication at times become tricky or difficult because of the language* barrier.' *(FG4-E5).*
	Theory-practice gap	Only the OSCE	*'We have the standard OSCEs that take place at the campus…' (Interview3-AS).*

Third theme **The OSCE quality and design**		Unobserved OSCE stations	*'In unobserved OSCE… We put the students in one place as each student has his separate PC and in one go over the computer (they complete the exam).' (FG1-D3).*
		Lack of individual feedback	*'…not individually. Individual written feedback could be unattainable or very difficult for the examiners...' (FG2-D3).*
	Reliability issues	Number of stations	*'We usually have from five to a maximum of six stations.' (Interview2-AS).*
		Parallel circuits and examiners' challenges	*'Last year, we had a very lengthy OSCE that, I think, was too much for the examiners. This is something that needs to be taken into account.' (FG4-E5).*
		Marking schemes design	*'We have a good rubric guiding the checklist, so non-specialist examiners can fairly conduct the exam, I think.' (FG4-E3).*
Fourth theme **Resources and infrastructure setup**	Human resources	Lack of workforce	*'I did not find suitable examiners for some stations; I cannot find outside examiners easily either.' (FG3-E2).*
		Training for stakeholders	*'I got the experience only by doing the actual exam; I have not received any specific training for the OSCE, no.' (FG3-E1).*
	Instructional and logistical resources	Educational resources	*'I think we have enough teaching time for basic clinical skills.' (FG2-D3).*
		Logistical resources	*'And you need to see what type of manikins and what instruments are available in your skills labs to design your station. Manikins' availability affects the station type…' (FG2-D1).*