Title	Development and evaluation of an assessment tool for identification of elements of situation awareness in OSCEs in undergraduate medical curricula
Author(s)	Fischer, Markus Andreas
Publication Date	2018-07-20
Publisher	NUI Galway
Item record	http://hdl.handle.net/10379/7424

Development and Evaluation of an Assessment Tool for Identification of Elements of Situation Awareness in OSCEs in Undergraduate Medical Curricula

Markus Andreas Fischer MSc

A thesis submitted for the degree of Doctor of Philosophy

Supervisors of Research:

Dr Thomas JB Kropmans PT PPT MSc PhD

Dr Kieran M Kennedy MB MICGP MSc MHSc MMedSc

Medical Informatics and Medical Education
School of Medicine
College of Medicine, Nursing and Health Sciences
National University of Ireland Galway

July 2018

Chair of the VIVA: Prof Fidelma Dunne

Internal Examiner: Prof Gerard Flaherty

External Examiner: Prof Mathieu Nendaz

Contents

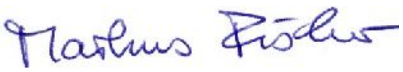
List of Tables

List of Figures

Declaration

This work has been submitted to fulfil the requirements of the degree of Doctor of Philosophy at the National University Ireland Galway. No part of this thesis has been previously submitted at this or any other university. Apart from due acknowledgements, it is entirely my own work.

Signed:



Markus Andreas Fischer

Date: 5th of July 2018

Acknowledgement

Personal experiences, combined with the recognised novelty of human factors as applied in healthcare, inspired me to commence this research project. While my life as mature student was full of privation over the last years, the overall learning experience, becoming acquainted with interesting field experts and the expansion of knowledge was immense and will contribute not only to my own mode of operation and working environment.

Now as the hard part is completed, it is time to express my gratitude to numerous clinical colleagues, co-research fellows and cooperative family members and friends.

First of all, my sincere gratitude appertains to Dr Thomas Kropmans. His dedication as principal supervisor and his ongoing impetus to complete the project was always the incentive for me to raise the bar towards the achievement of my desired goal.

Furthermore, I would like to express my sincere gratefulness to my co-supervisor Dr Kieran Kennedy for his insightful thoughts, valuable feedback and critical appraisal of my drafts prior to submission of the papers and the completion of the thesis.

The ongoing encouragement I received from both supervisors was the driving force for the expenditure of work I invested into the completion of the research and the thesis.

I would like to express my deep gratefulness to Prof Marlies P Schijven, Prof Stephen Durning and Prof Jean S Ker for their commitment as co-authors in reviewing the drafts, providing valuable feedback and offering recommendations for qualitative improvements. Their expertise within the field of medical education, clinical simulation and the assessment of clinical reasoning was very much appreciated.

Furthermore, I would like to express my gratitude to Dr Eva Doherty and Dr Paul O`Connor for their critical review of my papers and fruitful discussions regarding the association of aspects of cognitive psychology and clinical performance assessments. Their expertise in human factors applied in the healthcare environment provided guidance in the completion of this project.

Summing up, the specific expertise from different angles when applying methodologies to teach and assess cognitive information processing as part of undergraduate medical education supported the submission of high-quality papers for publication.

I am very grateful to the members of the Graduate Research Committee, Dr Eva Doherty, Dr Dara Byrne and Prof Andrew Murphy, for their critical appraisal of the progression of my research project and their substantive guidance throughout the years.

Furthermore, I would like to express my sincere appreciation to Prof Fidelma Dunne as Chair of the VIVA, Prof Gerard Flaherty as Internal Examiner and Prof Mathieu Nendaz as External Examiner at the rigorosum.

I also would like to thank Margaret Frere and John Tepper for their engagement in carrying out the pilot study, which led the foundation for the development of the consecutive guided training programme and subsequent evaluation study.

Furthermore, I would like to thank Annemarie Hennessy and Osama Zeyou for their precious support in actively participating in the research. Without their genuine interest and continuous enthusiasm, the completion of the project would not have been possible.

I would like to extend my thankfulness to all, my parents and siblings and their families and close friends, who always ensured that my good health still has priority.

Finally, I would like to express my sincere and deepest gratefulness to my wonderful partner Agnieszka. While she certainly suffered most from my continuous mental absence, her concern was always with my well-being. Thank you for taking such a good care of me over the last years.

Not to forget Heidi, our German shepard. Knowing that the look into her puppy dog eyes will always convince me to get up and grab some fresh air, this was always a most welcomed study break.

Abstract

Situation awareness (SA) is recognised as an important factor in patient safety and clinical decision making in conjunction with diagnostic and therapeutic reasoning. Missed diagnosis and treatment errors have been identified as intensely influenced by cognitive abilities of clinical physicians. Medical education aims to qualify physicians for diagnostic accuracy and the ability to select the most appropriate treatment. Research accentuated that medical students have little insight into cognitive processing and SA in clinical scenarios. Furthermore, literature does not suggest a methodology to quantify students´ cognitive processing in a clinical encounter. SA is described as having the proficiency to obtain awareness of the surrounding and to integrate this consciousness into the situational context and potential future development.

The reduction of fatal outcomes and critical incidents in high-risk environments such as aviation, was attributed to an acknowledgement of the impact of SA on task performance. Pilots undergo practical simulation exercises mirroring critical scenarios during their training and subsequent professional careers. Endsley`s model explains SA as an interdependent three-levelled concept, which enables one to distinguish the individual processes and the product of these processes when evaluating the cognitive abilities of examinees. The model has been used amongst multiple medical disciplines for training clinicians utilising simulation scenarios. However, the educational benefit has not achieved an overall integration of the concept of SA into medical curricula. Objective Structured Clinical Examinations (OSCEs) have been suggested to facilitate the assessment of certain elements of SA similar to high-risk simulation. Though, the underlying cognitive processing to obtain and maintain SA is difficult to measure and to quantify. Medical educators and clinical tutors might not be aware of essential steps in collecting, interpreting and incorporating necessary information during patient encounters.

The initial purpose of this thesis was to review the literature with a view to identifying whether levels of SA based on Endsley's model can be assessed utilising OSCEs during undergraduate medical training. Therefore, a systematic search was performed pertaining to SA and OSCEs, to identify studies published between January 1975 (first paper describing an OSCE) and February 2017, in peer reviewed international journals published in English. Selected databases were searched for papers that described the assessment of SA using OSCEs among undergraduate medical students. Findings suggested that whole-task OSCEs enable the evaluation of SA associated with clinical reasoning skills. Furthermore, if these assessments

address the levels of SA, these OSCEs can provide supportive feedback and strengthen educational measures associated with higher diagnostic accuracy and reasoning abilities. This suggests an early exposure of medical students to OSCEs mirroring clinical practice to evaluate and facilitate SA in clinical encounters. At the same time, it highlights the need for examiners and developers of these whole-task assessments to be able to understand the model of SA and to identify assessment criteria which can be assigned to the sequence and the product of the cognitive information processing.

Upon the identification of elements of OSCE assessments which can be assigned to the levels of SA, the next study explored the ability of clinical professionals to identify and characterise SA in OSCE guides and OSCE score sheets. Due to the unavailability of an instrument, a self−developed tool was used, thereby yielding an inferential measure of SA. The outcome of the study revealed a strong internal validity of the tool, however, only a moderate interrater reliability has been identified. In order to improve the level of interrater agreement, and thus the objectivity of the measurement tool, a consecutive guided training was developed. An expected outcome was the beneficial impact on assessors' ability to select parameters of OSCE assessments which can be assigned as process markers for SA or the level achieved in this type of assessment. In order to investigate underlying factors for any disagreement between raters rather than their subjectivity in evaluating OSCE forms, the Generalisability Theory (G-Theory) was suggested.

The evaluation study carried out upon completing of the training programme indicated the beneficial influence on the ability of raters to identify and categorise elements of SA within OSCE forms. The G-Theory revealed key facets for variance: OSCE score sheets, Levels of SA, Items embedded in the Levels of SA, Interaction between Forms and Levels and Forms and Items embedded within Levels. Thus, it was demonstrated that the consecutive guided training improved the identification of elements within OSCE assessments, which can be attributed to individual levels of SA. Outcomes of the G-Study highlighted the need to improve the readability of OSCE forms, suggested to be achieved by a clear description of the assessment markers for students` performance and behaviour.

Further research can contribute to fostering the development of educational and assessment strategies in undergraduate medical curricula for SA, and thus, improving patient safety.

Chapter 1. General Introduction

1.1. Introduction

Patient safety has gained tremendous public attention over the last decades underpinned by the continued release of reports of clinical malpractice [1]. Medical errors, such as flawed diagnostic and treatment decisions are among the most prevalent contributing factors to serious adverse consequences in the healthcare environment [2], responsible for an estimated 10% of hospitalisations in Europe [3]. It has been suggested that these errors are intensely personal and influenced by the physicians´ knowledge and cognitive abilities such as defective information processing and verification [4-6]. Clinical Reasoning (CR) as the underlying cognitive process in diagnostic and therapeutic decision making is directed by the situation and context of the patient's condition [7]. The ability for CR necessitates recognition and incorporation of multiple individual aspects of a patient, which enables the selection of the best treatment option in any given clinical presentation [8]. The accumulation of cognitive errors within CR has been suggested as predictive of the genesis of harmful events to the patient [9]. Erroneous decisions commonly occur due to numerous reasons and often result in harmful experiences for patients but also physicians. Deficient diagnostic accuracy due to undetected signs and symptoms of a patient's condition or the inappropriate selection of therapeutic interventions are often underlying catalysts for poor patient outcomes [10]. Retrospectively, diagnostic and therapeutic errors have been disregarded due to the missing understanding of their development and existence [11]. However, it has been shown that these deceptive mistakes are prone to flawed cognitive processing of essential information rather than system malfunctions [12].

Initiated by the book "To Err is Human – Building a Safer Health Care System" published as long ago as 1999, an awareness of how individual professionals can contribute to adverse events and outcomes in healthcare was born [2]. Numerous reports identified the human contribution to erroneous decisions in establishing diagnosis and treatment since then [4, 13, 14]. Factually flawed consequences include unnecessary diagnostic examinations potentially carrying harmful risk to the patient or potentially predictable and preventable adverse outcomes following medical interventions. The integration of best practices to remediate medication errors or to optimise efficient team performance resulted in the development of numerous methodologies for training and assessment in the healthcare environment [15-18]. Furthermore, decision-making tools and applications to support diagnostic accuracy and selection of best treatment modality incorporating evidence-based practice are increasingly described [19, 20]. However, research outcomes are pointing towards cognitive failures of the individual as a pivotal contribution when concluding erroneous clinical decisions [21]. Furthermore, various concomitants can facilitate the genesis of medical errors such as working conditions, environmental circumstances and system functions, but also the rapidly evolving progress in technological support solutions. Overall, the ultimate conclusion of patient care decisions is largely influenced by the clinicians' ability to reach a precise diagnosis and to subsequently select the most appropriate treatment and is suggested to correlate with the fundamental cognitive information processing as part of their CR [22]. Strategies for the development of CR abilities which can be embedded into medical training programmes have been studied over the years [23, 24]. However, errors within clinical reasoning practice continue to be reported, suggesting the critical appraisal of methodologies to foster efficient information processing which are successfully implemented in other high-risk environments [1, 5].

1.2. Rationale

1.2.1. Situation awareness

Situation awareness (SA) is thought of as a concept of cognitive processing of information in the field of human factors research [25-27]. However, no clarity of the composition of various underlying psychological processes exist due to the complexity in obtaining and maintaining SA [28]. Amongst 26 reviewed definitions and models, an even distribution referring to both SA as the cognitive process or SA as the cognitive state has been identified [29]. Further distinction has been described for the ways in which SA can be inferred from behaviour and verbalisation [30]. Implicit SA is commonly inferred from task-specific behaviour and actions when responding to a given set of circumstances. The fundamental cognitive processing of essential information required for task performance is suggested as the level of awareness of the situation. Thereby it is irrelevant if the participant is able to clearly verbalise the individual steps in gathering and utilising essential data. In contrast, explicit SA describes the conclusion of someone's consciousness regarding a situation which is inferred from behaviour but also from clearly formulated information. SA is commonly recognised as a condition, however, with low discriminant validity due to the similarity to concurrent mental constructs such as attention, vigilance, working and long-term memory and consciousness [31]. A generally accepted foundation of SA encompasses the allocation of cognitive resources towards elements essential for the completion of a task.

Since the identification and acknowledgment of the anthropogenic contribution on fatal outcomes in aviation accidents, the so called “Human Factors” are explicitly addressed in the aviation industry [26]. A mandatory training was developed as long ago as 1976 and continuously revised since, including non-technical skills, such as decision‑making and SA. The module includes aspects of the individual but also interactions of all available resources involved in completion of a given task [15]. Human limitations in recognition and understanding of a situation led to the introduction of SA training into the syllabus for pilots and crew members [32]. Based on the acknowledged beneficial impact of the delivery of this programme module on the reduction of fatal accidents or near miss incidents, the concept was adapted and is now in well-established use in numerous high-risk environments, such as nuclear power plants, oil platforms and rescue services [33]. Teaching and assessment of SA is embedded into practical simulation scenarios mirroring situations with potential critical outcomes. Numerous strategies have been developed to obtain and maintain SA, including proactively searching for and managing information, utilising checklists, and avoiding defective reasoning strategies [20].

1.2.2. Endsley`s model of situation awareness

In the widely acknowledged model developed by Endsley, SA is suggested to be a condition which interconnects process and state [34]. The syndetic three-level construct is best described in aviation as "A person´s mental model of the world around them" [26]. Endsley addressed that knowledge about a given set of actualities is central to effective decision making and ongoing assessment in dynamic systems [7, 34, 35]. The incorporation of the surrounding circumstances, the given set of actualities and their possible impact on future outcomes has been divided into three different levels of SA: Level 1 Perception, Level 2 Comprehension and Level 3 Projection (Figure 1.1) [34].

SA Level 1	Level 2 SA	Level 3 SA
Perception of situational elements	**Comprehension of elements in situation**	**Projection of their meaning for future situation**
Identification and collection of essential information from all available sources enabling to develop the “big picture” of what´s going on. The development of strategies to ensure the completeness of all relevant data is essential to understand the situation.	Assignment of meaning to the gathered information enabling to understand the situation and its criticality. Assessment of the integrity and accuracy of the data and their relevance to the situation supports the identification of information conflict or the necessity for expedient information.	“Thinking ahead” of optional outcomes of potential procedures or the planning of further essential activities supporting an accurate appreciation of future developments and the progression enabling the subsequent decision making process.

Figure 1.1: Levels of SA based on Endsley`s model [17]

These aspects are: Level 1 SA, accounting for the perception of the situational elements and parameters and is described as being responsible for 76% of errors in aviation caused by incorrect or missing gathering of relevant information. The primary underlying step is the collection of essential data about the situation. Characteristics of elements necessary to obtain the “big picture” of what is going on suggest the development of strategies to ensure the completeness of all data essential to achieve a comprehensive situational understanding. Information sources must be identified for all relevant information to allow subsequent task performance. Concurrently with focussing on the selective collection of data, additional gathering of unmentioned cues must also be noticed. Moreover, responsible for a further 20% contribution to errors, Level 2 SA describes the comprehension of the collected data and understanding of their relevance to the situation. If the multiple pieces of information cannot be put into relation, the situation is not understood. Therefore, the extracted information is utilised to develop a comprehensive “big picture” of the situation. Critical appraisal of the information supports the assessment of the integrity and accuracy of the data and their relevance to the situation. Finally, Level 3 SA describes the appreciation of the aggregation of collected data for their possible impact on the forthcoming progression of the situation. This includes the “thinking ahead” of optional outcomes when planning further essential activities. Identification of missing data to completely understand the given set of actualities and circumstances might transfer the cognitive process back to Level 1 SA when searching for more expedient information. In the next step, the additional data will be analysed again for its relevance to the situation. SA requires the ability to retain a vast amount of information obtained from numerous sources over a length of time [36]. Furthermore, the gathered data must be interpreted and structured in order to form a comprehensive understanding of the situation. The ability to integrate successive information and identify conflicting perceptions is an essential precondition for maintaining adequate SA [34, 37].

Endsley, in her model of SA for high-risk environments accentuated three interdependent levels essential to obtain and maintain awareness of the given situation and to project its possible development in the near future [26]. It is a model shown fit to describe the dynamic process of receiving, interpreting and processing information in dynamic environments such as the medical field [38, 39]. With regard to clinical practice, SA is believed to be essential for recognising and interpreting the clinical symptoms and signs of a patient`s illness, thereby enabling accurate CR [38, 40-42].

In healthcare, inadequate SA was identified as a primary parameter associated with deficient clinical performance, recommending the implementation of SA training including simulation into medical undergraduate education as realised in other high-risk environments [38, 43]. The WHO emphasised in 2009 the importance of early exposure of undergraduate medical students to elements of information processing to obtain as well as maintain SA [44, 45]. The WHO identified inadequate SA as a primary parameter associated with deficient clinical performance [44], recommending the implementation of “human factors” training as realised in other high-risk environments in medical undergraduate education [45]. Furthermore, SA was emphasised as one of four fundamental cornerstones incorporated in patient safety education into an undergraduate medical curriculum [46].

The recent implementation of mnemonics such as ISBAR (Identify, Situation, Background, Assessment and Recommendation) or I-PASS (Illness severity, Patient summary, Action list, Situation awareness & contingency and Synthesis by receiver) into healthcare highlights the importance of SA in improving safe and complete transfer of critical information [47, 48]. Learners seeking assistance from clinical experts are expected to provide appropriate and pivotal clinical information and observations based on the given presentation of the patient. On the basis of Endsley’s` model, elements of medical practice can be assigned to each of the three levels of SA.

Two aspects of Endsley´s model are widely debated which have implications for this thesis. The first predication suggests that SA enables one to control the steps of information gathering. The second proposition suggests that SA determines behaviour and activities of the operator in a situation. Endsley concedes the similarity between her model and the model of information processing, thus, her concept offers a useful framework for categorising elements of perception and the subsequent processing of important information. This is similar to Klein who concluded that the process of obtaining SA is guided by actively searching for information rather than a passive uptake of provided data [49]. Furthermore, he accentuated that SA is specified by the situation but also the response to the situation.

1.2.3. Assessment of situation awareness

Due to the missing consensus in the underlying concept of SA, there is also poor agreement on best assessment techniques [25, 50]. Currently applied methods can be divided into direct and indirect analysis [51]. Direct evaluation of the perception und utilisation of situational elements can be interrogated by both specific queries in a halted scenario or by a retrospective questionnaire upon completion of a simulation exercises. As an objective evaluation in real-time, one direct technique is described by the Situation Awareness Global Assessment Technique (SAGAT) [52]. Upon freezing the ongoing simulation exercise and blinding all sources providing relevant information, students can be probed for awareness and incorporation of situational elements required for adequate task performance. The Situational Awareness Rating Technique (SART) utilises post-exercise questionnaires and is a subjective approach to evaluate testees` self-estimated perception awareness of the particular situation.

Generally, it is suggested that the response to these situational queries sheds light on the understanding of the situation including the levels of SA based on the three tiers model. However, while the obtained level of SA can be determined, SAGAT does not provide insight into the internal cognitive processes itself. Indirect measuring techniques include the "think aloud" during task performance or the analysing of observed performance activities based on pre-determined behaviour or activity markers of participants in simulated scenarios. Trained assessors actively participating in these simulation scenarios can allusively direct the focus on missed opportunities to identify essential elements of SA.

In the Situation Presence Assessment Method (SPAM), participants of the simulation exercise are also queried specifically about the ongoing situation [53, 54]. In contrast to SAGAT, all relevant information is visible while the simulation is ongoing. On request, testees are probed for their data collection and analysis in a timely manner. Rather than relying on the memory capacity of the participant, SPAM tests the knowledge of where to extract and how to use essential data [54]. Therefore, the response time to provide correct answers and the accuracy of the information with regards to the content are utilised as quality markers for SA.

All techniques have the common requirement to facilitate the evaluation of the development and utilisation of SA in a given set of circumstances. For the ongoing improvement of these methods, a critical appraisal of behavioural and performance markers promoting a comprehensive understanding of the situation and the retrospective analysis of erroneous reasoning steps is essential.

1.2.4. Cognitive clinical competency

Cognition is characterised as the capability to employ knowledge, awareness of thought and information processing steps which include comprehension, inference, decision-making, planning and learning [55]. Furthermore, cognition facilitates the creation of the knowledge base and the establishment of concepts, which in medicine potentially enables the development of illness scripts in undergraduate students [56]. Competency-based medical education was introduced to teach and assess biomedical knowledge, clinical skills, and cognitive abilities [57]. However, assessment of clinical competency is a challenge due to the intricacy of the fundamental frameworks [58, 59], the defining of competency and the way it develops and how it can be purposefully assessed [60]. One definition suggests that competency is a complex construct in a dynamic process, rather than purely demonstrating knowledge and practical skills required by the profession of interest [61]. Furthermore, it has been stipulated, that competency includes the ability to integrate various subdomains of competencies when responding to a range of situational clinical challenges [60].

The development of clinical expertise is separated into four different levels (Figure 1.2) [62, 63]. Students, initially characterised as "unconsciously incompetent", learn clinically from experienced doctors who apply pattern recognition in their daily practice when assessing patients [64, 65]. Novices often are cognitively overburdened by the vast amount of available information and the prioritising process in identifying essential data, resulting in an incomplete or defective perception of the situation [66]. Professional clinicians who have developed their mental models by integration of knowledge and expertise over many years, are termed "unconsciously competent" [67]. The utilisation of illness scripts and schemata enables fast non-analytical thinking (System 1) resulting in an expeditious "big picture" of the clinical presentation of the patient, which is more comprehensive and projects possible outcomes when compared with the mental models of novices [66]. If the situation is not completely understood, clinical experts are able to switch to analytical thinking (System 2) [21]. However, they are commonly unaware of elements of SA and therefore, generally cannot convey or teach this sequence of data gathering and incorporation into the reasoning process [38, 68]. As a result, observing senior tutors might not enable students to develop incremental levels from conscious incompetence towards conscious competence through perceiving the essential steps of identifying and integrating relevant information for CR [67, 69]. Furthermore, Kiesewetter et al. emphasised, that very little knowledge exists about cognitive processing by medical students which may limit instruction on the incremental steps in CR in medical education [70].

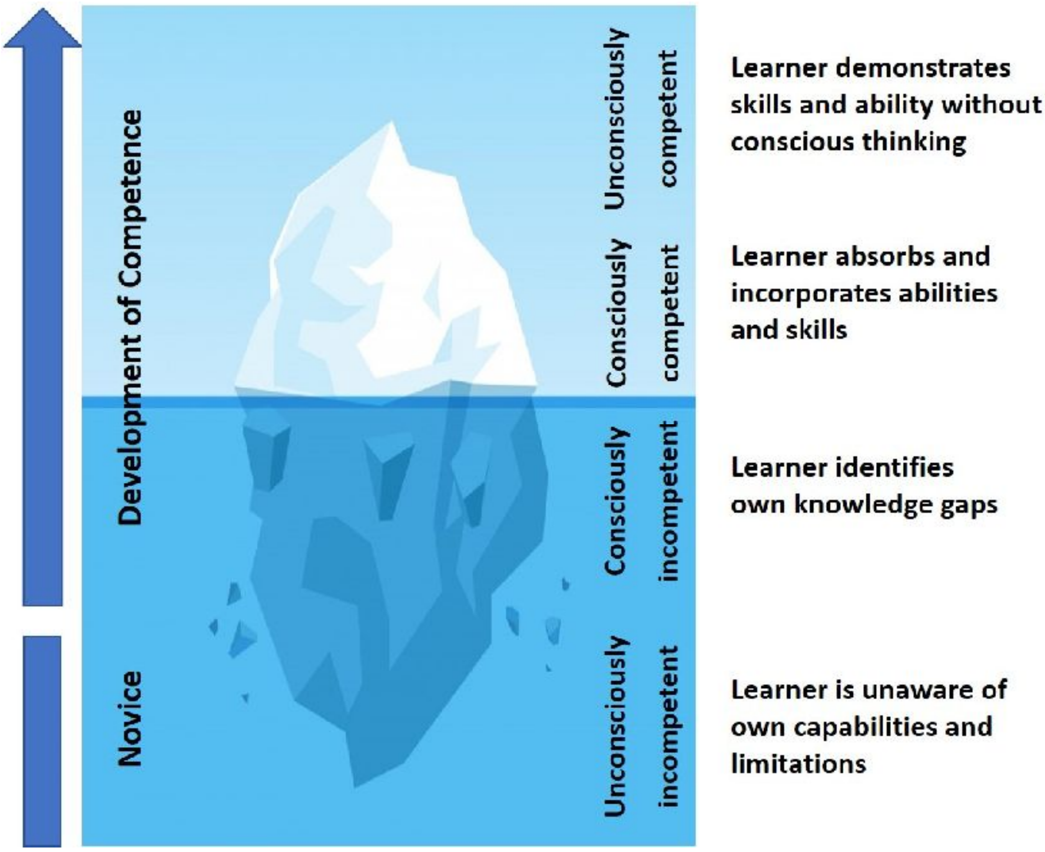


Figure 1.2: Developmental stages in competence according to Scott [71] (designed by Vvstudio - Freepik.com)

Twenty years ago, Goss highlighted the fact that medical students enter their third year of training competent in information gathering and facilitating patient care, but with deficient diagnostic reasoning ability [35]. Upon providing either a clinical vignette format or a chief complaint format in a paper-based examination, Nendaz and colleagues compared students, residents and general internists abilities in considering differential diagnosis (SA Level 2) or selecting basic diagnostic assessments (SA Level 1) and considering treatment options (SA Level 3). Thereby they noted that students were seen to be able to demonstrate knowledge and carry out examinations but struggled to incorporate the data into further diagnostic processes [72]. Because the utility of the data gathering process is closely linked with the process of subsequent reasoning, both should be jointly addressed and evaluated. More recently, Schuwirth argued that the outcome-based assessment does not reflect clinical reasoning abilities, and therefore, adequate alternative evaluation techniques of intermediate steps should be explored [73]. Singh et al. suggested a change in the current framework of the analytical diagnostic process in order to identify breakdowns in SA. By distinguishing the level at which SA was lacking, distinct measures can be applied in subsequent training [74]. This suggests the necessity of emphasising the understanding of SA in the medical context and of formulating novel potentials to teach and evaluate the utilisation of SA in educational healthcare settings.

There is little evidence that professionalism is a measurable construct based on psychometric data. Thus, it is suggested that competency cannot be directly evaluated or measured. A scoring system to assess competencies must therefore enable the examiner to infer the level of a particular or overall competency based on behaviour or activities. Clinical performance assessments demonstrate strong significance in medical curricula [75]. Qualitative assessment of CR in medical students with a high expressiveness could potentially allow the inference of students` cognitive competency in information processing. Furthermore, the opportunity for providing direct and meaningful feedback based on the assessment offers a learning opportunity for both; the medical faculty and the student. However, the development of assessment protocols and performance score sheets remain a challenge for tutors and educators in healthcare settings. While a stabile framework of biomedical knowledge is still the fundamental prerequisite for medical trainees, additional emphasis must focus on the requirement for cognitive competencies in information processing [76]. It has been suggested that identifying competencies, representing expertise in diagnosis and selection of therapies by medical students, will support their attainment of professionalism in medical care.

1.2.5. Undergraduate medical education

Medical education aims to qualify physicians in both diagnostic accuracy and subsequently selecting the best treatment option for any given patient presentation [77]. Notwithstanding the implementation of innovative teaching and assessment methods, such as simulation-based learning [78, 79] and problem-based learning [80, 81] into medical education curricula, flawed identification of the clinical presentation and defective appropriateness of therapeutic options continue to be reported [82-84]. In medicine, SA is essential for recognising the clinical symptoms and signs of a patient´s illness, thereby enabling the clinician to apply adequate clinical reasoning strategies with respect to diagnosis and best treatment options [38, 40, 43]. However, the positive outcomes of the human factors training seen in high-risk environments are merely acknowledged in healthcare and have only recently been embedded by certain medical disciplines and faculties [85, 86]. However, due to the increasing number of reports of diagnostic errors and false treatment decisions including prescribing errors, the recognition of the training adapted to the specific requirements of the healthcare environment can be anticipated.

Clinical Reasoning (CR) includes the fundamental cognitive information processing to reach diagnostic and therapeutic decisions and has been shown to be governed by both the patient`s signs and symptoms and the situation in which they occur. Acquiring CR presupposes both the ability to identify the underlying causes for a patient`s condition as well as the ability to extract and integrate additional information needed to fully understand the clinical situation [87]. Deficient information processing of physicians is reported throughout the literature, suggesting the exigency to develop strategies to foster more competent cognitive reasoning abilities [1, 5]. Ongoing research in the field of diagnostic reasoning and clinical errors is mainly carried out retrospectively. To date, not much evidence exists on the issue in professional healthcare settings such as primary care or speciality training [14]. Nevertheless, findings within these clinical areas support the identification of subdomains within the CR process that contribute to erroneous consequences. Conclusions of these studies are pointing towards the necessity to develop methods to assess the clinicians` cognitive ability for diagnostic reasoning [63, 88, 89]. Defective CR resulting in erroneous diagnostic and harmful therapeutic consequences for the patient can be the outcome of poor organisation of biomedical knowledge and data interpretation or the combination of both [63, 72]. The ability for CR is regarded as a pivotal measure of competency and, thus, should be explicitly taught and assessed at an early phase of medical training.

Clinical experts can select illness scripts, developed over time during clinical practice, when diagnosing a patient's condition. In contrast, undergraduate medical students are only conceptualising their knowledge and developing illness scripts [67]. Thus, accurate synthesis of signs and symptoms of the clinical presentation of the patient to reach the definite diagnosis has not been established. Medical educators and clinical tutors have little insight into the framework of information organisation and retrieval of already acquired knowledge and incorporation of gathered data by medical students [68]. Therefore, there is a necessity to understand how scholars structure their biomedical knowledge and integrate perceived information from various sources into their daily clinical practice. A basic understanding of the process could facilitate the delivery of educational methods in developing meaningful connections to reach clinical diagnosis and select therapeutic options. The World Health Organisation recommended in their guidelines in 2009 the implementation of teaching methods similar to aviation into undergraduate medical education [44].

Furthermore, research outcomes direct the focus on the development of educational strategies which can be implemented into early medical training including assessments such as the Objective Structured Clinical Examination (OSCE) [24]. The OSCE is a well-established type of examination [90] often used to assess performance in the medical domain [91]. It is designed to test clinical skill performance and competence across multiple domains including communication, clinical examination and interpretation of results [92].

1.2.6. OSCE assessments

OSCEs are utilised internationally as an assessment tool for knowledge, skills and clinical performance [90, 93]. In addition, OSCEs are, in theory, intended to function as educational measures during medical training allowing for the assessment of students` competence under variable circumstances [93-95]. Fida and Kassab showed that scores achieved by medical students in OSCE stations demonstrated strong predictive value for the students´ ability to identify and integrate relevant information and competently manage a patient [8]. Therefore, there is potential for the identification and remediation of deficits in selecting and integrating essential parameters, which is pivotal for CR [65]. Thus, OSCEs allow for the evaluation of the substance and effectiveness of the curriculum. Utilisation of simulated patients and time-restricted identification of important parameters during OSCEs mirrors the clinical practice environment, in which assessment of the situation and subsequent CR occurs under time and workload pressure [94]. Few studies have investigated the efficacy of an OSCE for evaluating medical students' CR ability. Contrary to that, Martin et al. demonstrated no significant correlation between OSCE scores, data interpretation and CR [96]. In addition, research outcomes highlighted uncertainty as to whether scores achieved in OSCEs correlate with students` information processing [8, 96]. This suggests the necessity to develop a methodology to improve the evaluation of students` ability for applying essential steps to reach diagnostic accuracy and adequate selection of therapeutic options using this type of assessment.

The shift from time-based education to competency-based training in medicine necessitates the development of adequate assessment methods [97]. History taking and physical examination are core skills demonstrated by medical students. However, the ability to integrate the gathered information into further processing steps is a fundamental requirement for CR [59, 98]. Assessment in undergraduate medical curricula rarely incorporate cognitive information processing indicating the development and utilisation of SA embedded in the underlying CR process. Furthermore, evaluating professional skills based on human judgment of behavioural markers of testees are prone to subjectivity of the raters [99]. Attempts to mitigate this individual impact were strengthened by standardising the assessment and how the level of performance can be determined [100].

Whole-task OSCEs including elements of all three levels of SA have been suggested to inspire students to develop cognitive abilities to obtain diagnostic and therapeutic accuracy [101]. OSCEs do not provide a comprehensive evaluation of an overall competency [58]. However, if set up as summative evaluation, this type of assessment is suggested to draw an informative compilation of the students' ability to integrate various competencies [100, 102, 103]. Thus, OSCEs might be a suitable instrument to evaluate students' understanding of the situation as part of their CR and subsequently to provide deductive feedback of their cognitive processing upon completion of the patient encounter [104]. However, literature does not suggest an accepted methodology to quantify students´ utilisation of SA in a clinical encounter. Simply assuming that accurate SA automatically matches reasonable performance and vice versa has been disproven [105].

These factors raise the question of whether aviation-like SA training and assessment could be purposefully reflected in undergraduate medical education and assessment. OSCEs may be a suitable instrument to teach and evaluate students' use of SA as part of their CR.

Chapter 2 Literature review

2.1. Method

A systematic search of the literature was performed pertaining to SA and OSCEs, to identify studies published between January 1975 (first paper describing an OSCE) and February 2017, in peer reviewed international journals published in English (Table 2.1). PUBMED, EMBASE, PsycINFO Ovid and SCOPUS were searched for papers that described the assessment of CR using OSCEs among undergraduate medical students. Key search terms included “Objective Structured Clinical Examination”, Objective Structured Clinical Assessment” or “OSCE” and “non-technical skills”, “sense-making”, “clinical reasoning”, “perception”, “comprehension”, “projection”, "situation awareness”, “situational awareness” and “situation assessment". Boolean operators (AND, OR) were used as conjunctions to narrow the search strategy, resulting in the limitation of papers relevant to the research interest. Publications relating to undergraduate medical training and 'situation awareness' or information processing as part of clinical reasoning were included. Due to different cognitive demands and scopes of practice, publications relating to nursing, paramedical disciplines, pharmacy and veterinary education were excluded from the search. The abstracts of remaining papers were manually reviewed in order to ensure their relevance (Figure 2.4). Areas of particular interest were elements of SA within OSCEs and the assessment of SA within these examinations. Additionally, a manual review of the references listed in the remaining publications was carried out and any publications of potential interest were sourced and reviewed.

Table 2.1: Steps of initial literature search to retrieve papers for the critical appraisal of their relevance to SA and OSCEs in undergraduate medical education

1	Objective Structured Clinical Examination
2	OSCE
3	OR 1-2
4	Objective Structured Clinical Assessment
5	OR 3-4
6	Non-technical skills
7	AND 5-6
8	Sense-making
9	AND 5-8
10	Clinical reasoning
11	AND 5-10
12	Perception
13	Comprehension
14	Projection
15	OR 12-13-14
16	AND 5-15
17	Situation awareness
18	Situational awareness
19	Situation assessment
20	OR 16-17-18
21	AND 5-20

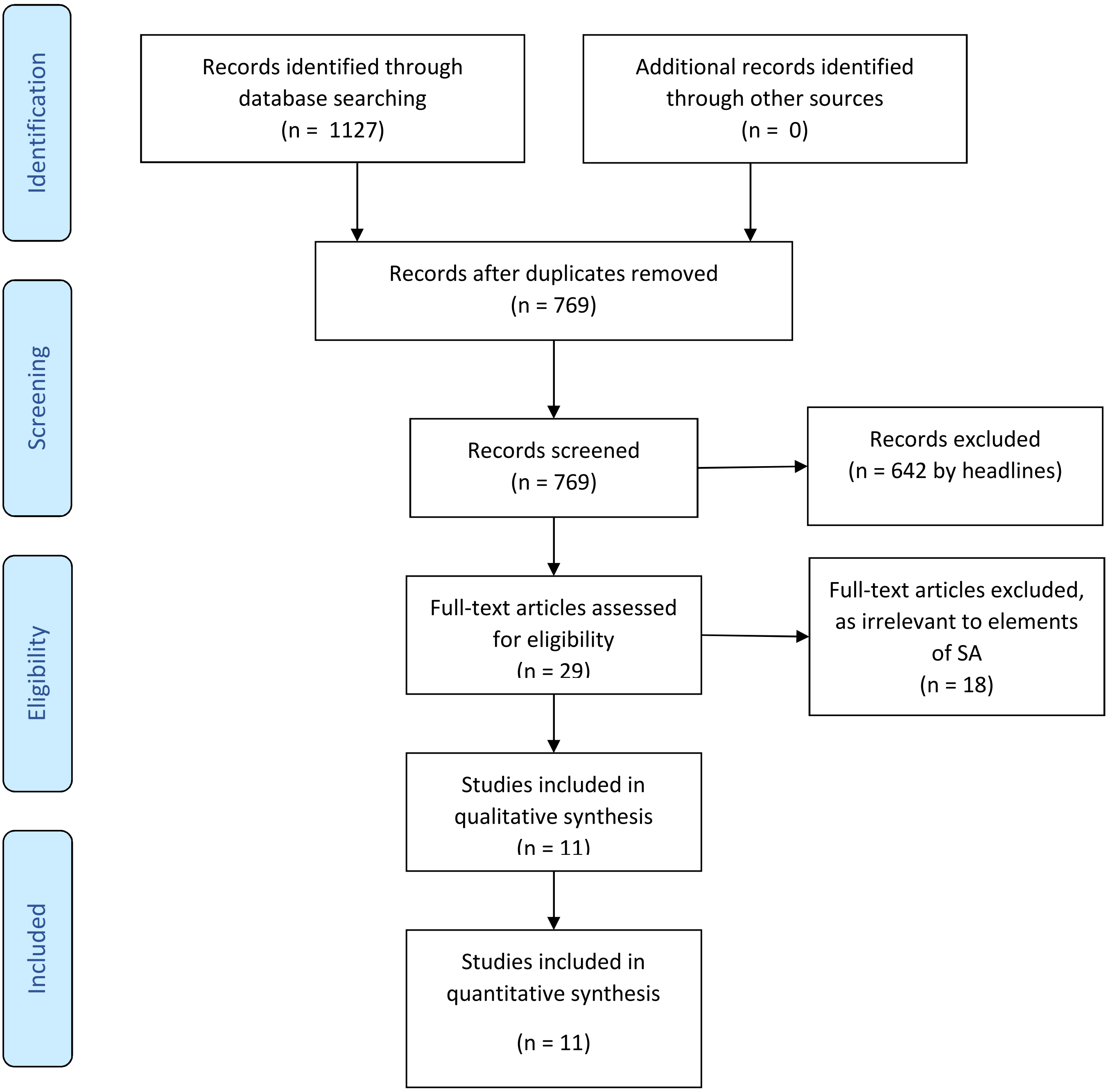


Figure 2.1: Flow chart describing the selection process (PRISMA)

2.2. Results

The search of the literature retrieved 11 articles eligible for inclusion (Table 2.2). Only one publication demonstrated an association between the OSCE and SA. An appraisal of the study design of the utilised simulation scenario, however, revealed that a root cause analysis was undertaken by the medical students to identify a prescription error [106]. Part of the examination focused on SA Level 1 when students were asked to take a history of the incident and SA Level 2 when integrating these data into the understanding of the situation. The authors suggested OSCEs to reflect utilisation of SA, however, neither a definition of the meaning nor the model of SA used for the conclusion was provided. Evaluation of SA Level 1 was identified in 11 publications, mostly seen in elements such as physical examinations, history taking but also in obtaining an overall impression of the patient and the retrieval of diagnostic test results. All 11 studies demonstrated consecutive evaluation of elements of SA Level 2, demonstrated by the integration of the gathered parameters in SA Level 1 into further information processing steps. Only two studies assessed the selection process of optional diagnostic and treatment modalities categorised in SA Level 3.

Six papers described the OSCE as having the potential to be an assessment tool for CR [107-112], a method that might correspond with those used for the assessment of SA in high-risk environments or simulation scenarios.

Furthermore, five papers suggested the OSCE as a valuable means for educating medical students on information gathering and processing when they are assessing the identification of the clinical presentation and incorporating the findings into their decision tree [106, 113-116].

Table 2.2: Results of the analysis of 11 identified papers relating to SA (SA Level 1,2,3) in undergraduate medical training as evaluated by OSCEs

Author / Year of Publication	Year of Study	Number of Students	Level of Education	SA Level 1	SA Level 2	SA Level 3	Feedback	Assessment Tool for SA	Educational Tool for SA	Research Interest
Volkan 2004	1999	169	year three	History taking, physical examination	Differential diagnosis	Consideration of treatment options		X		Factor analysis of OSCE constructs
Durak 2007	2000-2001	382	year six	Overall impression, history taking, diagnostic test results	Differential diagnosis	Consideration of treatment options, identification the need for further investigations	X		X	Case-based stationary examination
Varkey 2007	2003	42	year three	History taking	Identification of root cause of error		X		X	Root-cause analysis of error
Durning 2012	2010	170	year two	History taking, physical examination	Differential diagnosis			X		Feasibility, reliability, and validity of the evaluation of clinical reasoning utilising OSCEs
Myung 2013	2011	145	year four	Physical examination	Differential diagnosis				X	OSCE evaluation impact of pre-encounter analytical reasoning training
Lafleur 2015	2013	40	year five	Physical examination	Diagnostic reasoning				X	Influence of OSCE design on diagnostic reasoning
LaRochelle 2015	2009-2011	514	year four	History taking, physical examination	Clinical reasoning			X		Impact of pre-clerkship clinical reasoning training
Park 2015	2011	65	year four	Overall impression, history taking, physical examination	Differential diagnosis			X		Comparison of clinical reasoning scores and diagnostic accuracy
Sim 2015	2013	185	year five	History taking, physical examination	Data interpretation, clinical reasoning			X		Assessment of different clinical skills using OSCE
Stansfield 2016	2012	45	year four	Physical examination	Diagnostic reasoning			X		Evaluation of embedding clinical examination results into diagnostic reasoning
Furmedge 2016	2013/ 2014	1280	year one/ two	Information gathering	Predefined focus on integration of basic and clinical science		X		X	Acceptability and educational impact of OSCEs in early years

2.2.1. Situation awareness as part of the evaluation of clinical reasoning

Six studies concluded that OSCE stations allow for the assessment of students' utilisation of CR abilities within diagnostic thinking [107-112]. In a study by Durning et al. based on three successive stations, students were asked to take a history from a patient, synthesise the data and provide the most likely diagnosis and a problem list. In the last step, the patient had to be presented to an attending colleague [107]. La Rochelle and colleagues detected a correlation between clinical and reasoning skills during pre-clerkship and abilities observed during internship [108]. Therefore, they suggested the potential of OSCEs to identify and foster those students who are experiencing difficulties with diagnostic reasoning and so possibly to prevent problems in subsequent clinical performance. Park et al., in contrast, demonstrated the inability of OSCE scores to correlate with CR abilities [109]. However, they demonstrated that scores achieved in CR OSCEs strongly correlated with diagnostic accuracy.

When assessing students across 16 OSCE stations, Sim et al. demonstrated that out of six evaluation criteria [history taking, physical examination, communication skills, clinical reasoning skills, procedural skills, professionalism] procedural skills were identified as strongest and CR abilities as weakest [110]. They suggested that the low mean scores could be the result of students` lack of biomedical knowledge, their inability to incorporate the collected information into the clinical presentation of the patient or a combination of both.

Volkan et al. in their study suggested two fundamental structures for OSCEs. Information gathering was represented by history-taking and physical examination, whereas reasoning and dissemination included hypothetico-deductive testing and differential diagnostic thinking [111]. Based on the findings of previous studies in which students showed a drop in clinical reasoning when focusing on history-taking and physical examination, they highlighted the importance of comprehensive OSCEs to assess the ability to apply both processes simultaneously. In an innovative OSCE assessing the connotation of CR and physical examination abilities, Stansfield and colleagues identified a discrepancy between integrating acquired knowledge into the selected physical manoeuvres [112]. Additionally, there were fewer deficits in employing adequate physical examination skills in students able to embed their findings into the CR process.

2.2.2. The OSCE as an educational tool for situation awareness

Five research groups identified the potential for OSCE stations to be teaching tools for SA within medical education [106, 113-115]. Generally, studies demonstrated better diagnostic accuracy and reasoning abilities among students when using an underlying analytical approach. Direct feedback or the addition of supportive information between incremental OSCE scenarios exemplified good educational properties. Durak et al. described a model in which hybrid forms of OSCE stations were applied [113]. Based on patient scenarios, students were asked to develop a treatment plan and were guided in a stepwise manner. The initial step included the collection of relevant data from history-taking, evaluating signs and symptoms, and the identification of underlying pathophysiological changes. After identifying the most likely diagnosis, students were probed to extract relevant information from the clinical notes and diagnostic results. Subsequently, students created the treatment plan for the patient based on the chosen diagnosis. In between these steps, corrective feedback was provided and incorporated into subsequent decision making. This method was found to be a motivator for students to improve their clinical reasoning.

Lafleur et al. observed the impact of the design of OSCE stations on the learning behaviour of students [114]. They described students applying more diagnostic reasoning when studying for whole task OSCEs rather than those that focused purely on physical examinations. Backward and forward associations, that is, either looking for evidence to support a suspected diagnosis or the aggregation of all identified symptoms and signs to conclude a diagnosis respectively, are both tasks that demand higher cognitive processing activities and, were strengthened when studying collaboratively for comprehensive OSCEs. Myung et al. compared analytical reasoning ability and diagnostic accuracy in a randomised controlled study [115]. On analysis of two groups of students, one of which had received prior education on analytical reasoning and one of which had not, OSCE scores achieved in both cohorts demonstrated no difference for information gathering. However, higher diagnostic accuracy was seen in that group of students which had received training in applying analytical reasoning strategies.

Due to the similarity to real clinical situations, Varkey et al. suggest that OSCEs in general are an ideal tool for assessing and teaching SA [106]. However, no statement of the meaning of SA or the association with the healthcare environment was provided. In their study, students were asked to identify pivotal information in an error-induced patient encounter. Formative feedback was provided by the tutor on information gathering, root cause analysis, and completing the task. Furmedge and colleagues interrogated the appreciation of students for a novel, formative OSCE. The clinical scenario was designed to enable testees to exemplify the integration of skills and knowledge into the understanding of a situation rather than the pure retrieval of recited text passages. In this study, OSCEs were seen as a learning environment to develop cognitive strategies when exposed to clinical scenarios mirroring reality [116].

2.3. Discussion

We suggest that OSCE stations could be utilised for the assessment of elements of SA in medical students, using whole task simulation scenarios. So far, no distinct comprehensible methodology has been described which is universally accepted as a fundamental measurement of SA. Furthermore, the conjecture that accurate SA automatically correlates with adequate performance and vice versa has been disproven [105]. Although students may demonstrate history-taking, physical examination and procedural skills, the literature suggests that they are frequently unable to embed their findings in subsequent steps and decisions. This might be explained by the fact that novices often only recite enormous amounts of information from their "knowledge database".

Reduced diagnostic accuracy by medical students accentuated the primary necessity for efficient data gathering and processing [63, 72]. Diagnostic excellence has been suggested to originate from a reasonable understanding of the fundamental anatomical and physiological context in conjunction with pathophysiological changes potentially identifiable within elements of SA in any given clinical presentation [117]. Borleffs et al. described the objective of teaching CR as the ability to make correct decisions in the process of establishing a diagnosis [118]. Alexander concluded that students must be able to demonstrate how to do it, but also, at the same time, why to do it [119]. Zwaan et al. suggested implementing interventions with proven records to enhance SA within the diagnostic reasoning process [9]. Gruppen and colleagues depicted how the different utilisation of hypotheses and information depends on clinical experience and expertise [120]. In their study, the collection and appropriate selection of data was demonstrated to be more difficult than the pure integration of available information. This imbalance between efficient information gathering and successive data integration suggests that educational measures should aim to enhance procedures in collecting and processing relevant information.

2.3.1. The OSCE as a learning approach for SA for medical students

OSCE stations can be educational tools for CR, pattern recognition and problem-based learning [121]. To foster the ability of putting it all together, Furmedge et al. suggested an early exposure of students to OSCEs [116]. However, they concurrently highlighted the need to identify how early OSCE exposure could contribute to development of non-analytical reasoning skills. When analysing feedback upon completion of the OSCE cycle, Haider and colleagues summarised students` appreciation of this type of assessment, which supported their individual abilities to identify areas of clinical weakness, thus inspiring their interest in developing information processing skills [122]. Baker et al. introduced three strategies for developing CR, hypothesis testing, forward thinking and pattern recognition [123]. They developed a specific assessment tool for the interpretative summary, differential diagnosis, explanation of reasoning and alternative diagnostics [IDEA].

OSCEs were described as a means of valuable feedback for both, examinee and educator [124], that enables the reinforcement of the importance of SA as an underlying requirement for well-informed CR in all disciplines [37, 63]. Feedback provided upon completion of OSCE scenarios could support the faculty's appraisal and the examinees' self-rating of the sense-making process when selecting best clinical diagnosis and therapeutic options [113]. Providing individualised feedback upon completion of the OSCE was described as being complex [125]. Thus, establishing the cognitive map of the underlying information processing could potentially identify why selected parameters and criteria during the CR process either made sense to the testee at the time or were neglected [126-128].

Remedial teaching and education at undergraduate level could be considered if a deficiency within the three levels of SA was identified during OSCE assessments [129]. Gregory et al. described an innovative method of teaching aspects of SA in undergraduate medical training by exposing students not only to perils, but also to additional indications of a patient's condition [130]. Upon entry into undergraduate training, students are exposed to a clinical area without a patient, such as the bed space, and are evaluated collectively in their ability to recognise any hazards and clues indicating supportive information about the clinical status of the patient. Students are also expected to extract additional parameters from clinical notes and diagnostic results. The positive feedback from students and tutors suggests that this approach is a promising tool in teaching SA to medical students.

2.4. Conclusion

Assessment of elements of SA as adapted from the model by Endsley might have the potential to be translated into certain aspects of CR evaluation using OSCEs. Given that assessment is a fundamental driver of adult learning, incorporating the quantitation of utilisation of SA within OSCEs during undergraduate medical training could develop and strengthen teaching on information gathering and efficient processing. However, further research needs to establish whether different levels of SA can be identified throughout the medical curriculum and its assessment including the use of paper cases and reviewing medical records. If so, are these levels of assessment congruent with the learning outcomes in preclinical and clinical years? In order to teach students how to perceive and incorporate relevant data, it is essential to provide focused and informative feedback related to each level of SA and the associated steps of CR.

Upon identification of the potential and ability to assess levels of SA in a curriculum eg. OSCEs, we suggest that students be exposed, in a staged format, to the concept of SA at the early stages in their training, prior to meeting complex challenging clinical situations in their later medical careers. Efforts in conveying underlying elements of SA during undergraduate education could be reflected in enhanced abilities to read and understand clinical scenarios in subsequent clinical practice.

Chapter 3: Study about assessing situation awareness in medical education

3.1. Rationale for the study

Medical educators call for both instructional methodologies and assessment techniques to convey the importance of obtaining the “big picture” of an ongoing situation. The potential to identify if SA was obtained and utilised by medical students or where the chain of SA was broken during clinical performance measures is suggested to support the development of professional expertise. Medical educators and tutors have little insight into the information interpretation and incorporation during patient encounters by undergraduate medical and nursing students which hinders the evaluation and teaching of the underlying cognitive processing during clinical assessments [131]. Students commonly demonstrate knowledge and the ability to gather relevant information during examinations, however, struggle to incorporate this gathered data into further cognitive proceedings. The imbalance between efficient information gathering and successive data integration suggested that educational measures should aim to enhance procedures in collecting and processing relevant information [132].

The recent implementation of mnemonics such as ISBAR (Identify, Situation, Background, Assessment and Recommendation) or I-PASS (Illness severity, Patient summary, Action list, Situation awareness & contingency and Synthesis by receiver) into various medical disciplines accentuates the significance of SA for the safe and comprehensive transfer of critical information [47, 48]. Novices asking for advice or instructions from clinical experts need to be able to provide pivotal aspects and observations of the patient’s clinical presentation. To foster the ability of "putting it all together", an early exposure of students to Objective Structured Clinical Examination (OSCE) was suggested [116]. However, the need to identify how this type of assessment could contribute to development of cognitive abilities he concurrently highlighted. Scores achieved by medical students in OSCE stations demonstrated strong predictive value for their ability to select and integrate essential information and subsequently to develop the therapeutic plan for the patient [8].

Performance and clinical skills assessments of healthcare students rarely incorporate cognitive elements [75]. Furthermore, experienced practitioners developed their expertise in integrating elements of the patient encounter to conclude subsequent diagnosis and therapy over many years [68]. Clinical tutors and medical educators are commonly unaware of the essential steps of the fundamental information processing. Thus, they might not be able to communicate or evaluate the underlying cognitive process of information gathering and utilisation during clinical assessments. This highlights the necessity for a clear understanding of the model of SA and its potential application within the healthcare environment by clinical tutors and developers of OSCE stations. Wilkinson et al. demonstrated a correlation between direct involvement of the examiners in the development of OSCEs including objectives, format and score sheets and associated improvement in the examiners understanding [133].

The usability of a novel assessment tool has yet to be proven for its reliability and validity in order to gain acceptance by medical faculties. Therefore, the potential existence of elements of clinical practice which are assessed in OSCEs needs to be evaluated for their property to be categorised to the individual levels of SA. In addition, acceptability of a novel tool needs a proven record of its desired outcome measure. The involvement of educators and tutors in the development of OSCEs including the assessment of SA is essential for the elaboration of the fundamental concept of cognitive information processing. Thus, a further requirement for the practical use of an assessment tool is the ability of medical educators and clinical tutors to identify elements within the OSCE scoring system which can be assigned to the individual levels of SA. A collaboration with medical educators, clinical tutors and human factors psychologists from different faculties was part of this research project.

3.2 General aim

The general aim of this research project is to identify the feasibility of the OSCE to teach and evaluate the utilisation of SA by undergraduate medical students based on the model developed by Endsley. The research questions have been developed in order to support the general aim of the study. Diagnostic and treatment errors are suggested to reflect both the physicians´ incomplete biomedical knowledge base or impaired cognitive abilities such as defective information processing and verification or both [4-6]. CR includes cognitive processes in diagnostic and therapeutic decision making and has been shown to be directed by the situational context of the patient's clinical presentation and circumstances [9]. The potential for OSCEs to develop a cognitive map of the students` underlying CR steps could facilitate the development of methodologies to teach on information gathering and efficient processing. However, in order to convey how to perceive and incorporate relevant data, it is essential to provide focused and informative feedback related to each level of SA and the associated steps of CR. Therefore, there is a need to identify whether medical educators and clinical tutors are able to categorise essential steps in information processing to the levels of SA.

3.3. Purpose of study

The purpose of the pilot study was first to determine validity and next to determine interrater reliability of a SA assessment tool and to determine the degree of SA present in several medical student OSCE guides.

The purpose of the subsequent study was 1. to evaluate the effect of consecutive 'guided' training on the improvement of interrater agreement, 2. to assess the reliability of a method for identifying elements of SA embedded in Objective Structured Clinical Examination (OSCE) station score sheets, which can be categorised to the three levels of SA (Level 1 SA, Level 2 SA, Level 3 SA) based on Endsley's model, and 3. to identify facets contributing to interrater disagreement. The validity and reliability of OSCE assessments is pivotal in order to evaluate students' competence in the selected domain. Poor reliability hinders the ability of the faculty to examine the performance in these types of examinations.

The main research outcomes are interpreted with reference to the aforementioned research questions. Suggestions for further research in how best to teach and assess SA in undergraduate medical curricula are provided based on the implications of these research findings.

Chapter 4: Methods

4.1. Development of Assessment Tool

Initially, a phased course of action was developed (Figure 4.1). In phase 1, elements extracted from papers identifying underlying causes of diagnostic and treatment errors in clinical practice were classified to the individual levels of SA in our model used. [14, 120, 134]. S Subsequently, this information was critically appraised and elements categorised to facets (Table 4.1). Causative elements of the clinical practice contributing to impaired information gathering and integrating into subsequent cognitive processing were identified. In phase 2, an assessment tool was developed based on the collected data in phase 1, aiming to enable educators to evaluate elements of SA. Essential steps during a patient encounter were identified and incorporated into a hierarchical diagram (Figure 4.2) in the style of Endsley's model. This tool was subsequently utilised to develop the node tree utilised for coding of OSCE forms with NVIVO. The Levels of SA demonstrate the parental nodes and facets amongst each individual level of SA the associated child nodes.

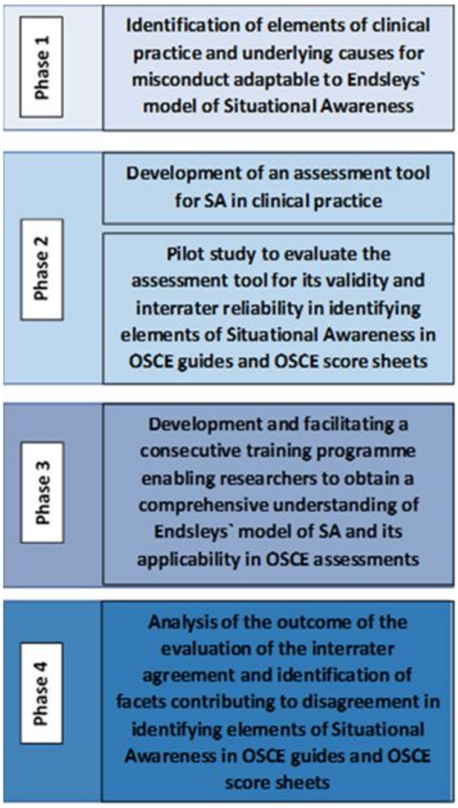


Figure 4.1: Phased course of action model

Table 4.1: Elements of clinical practice categorised according to the individual levels of SA based on Endsley's model

Level 1 SA: Perception of situational elements

Overall general impression	Diagnostic impression based on the clinical appearance of the patient
	Environmental scan including items indicating medical impairment of the patient
History taking	Chief complaint or reason for consultation History of course of present disease/ illness History of diseases independent from present illness Treatment and drug/ medication history
Physical examination	General physical condition and specific physical discomfort
Retrieval of diagnostic test results	Results of common and organ specific diagnostic laboratory tests Findings from diagnostic imaging

Level 2 SA: Comprehension of elements in situation

Pattern recognition	Recognition of concurrence of clinical signs, symptoms and/ or complaints
Detection of abnormalities	Recognition of unusual and/ or unsuspected findings and pathological changes Identification of information conflict Possible misinterpretation of complaints, signs and symptoms
Formulating working diagnosis	Determination of the most favourable disease based on the clinical presentation and gathered information
Consideration of differential diagnoses	Incorporating all gathered information into critical consideration of optional matching diseases

Level 3 SA: Projection of their meaning for future situation

Consideration of treatment options	Availability and restrictions of treatments options including side effects with ongoing treatment Harmonisation of the patients and physicians' preferences
Identification of need for further investigations	Necessity, reliability and validity of additional examinations and tests
Consideration of optional outcomes	Potential consequences (benefits and harm) harm of optional therapies and additional examinations and tests
Search for expedient additional information	Identification of absence of potentially valuable information Outlook for the patient

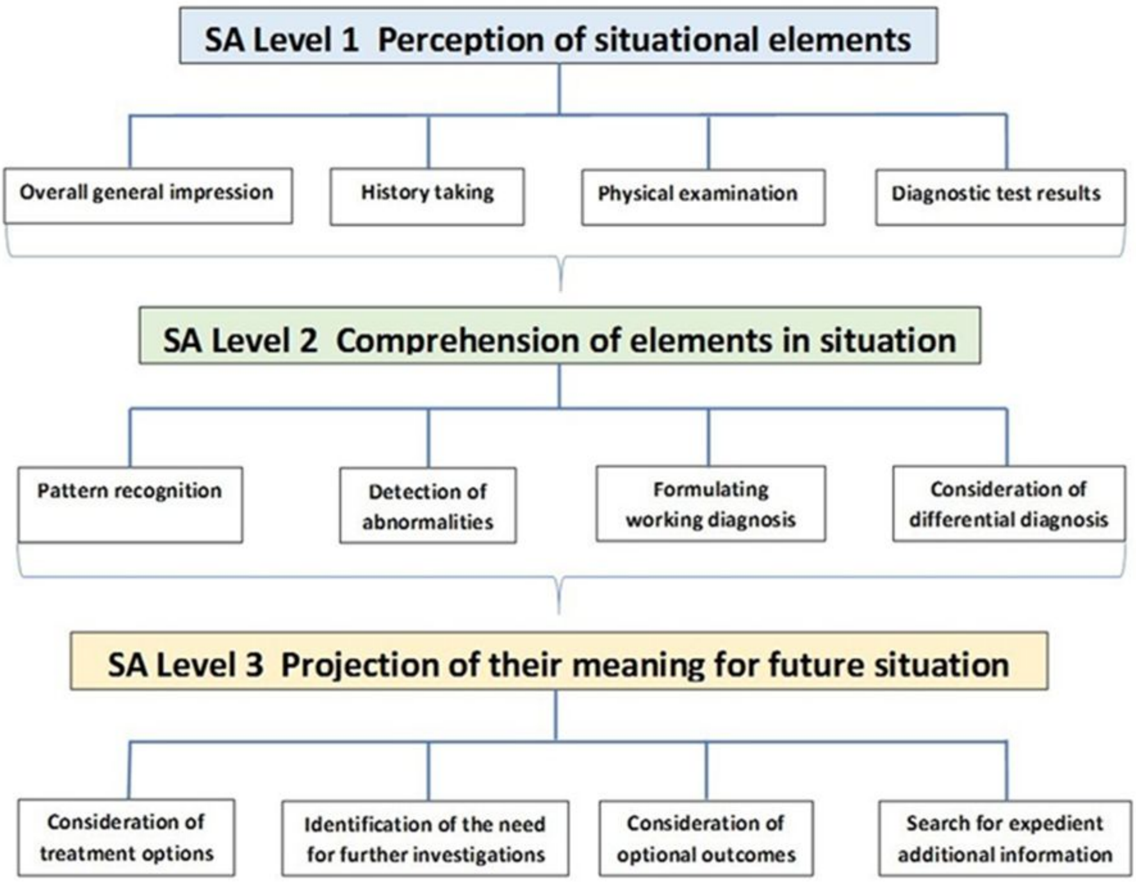


Figure 4.2: Self-developed tool based on Endsley's model of SA used to assess presence of SA in OSCE stations

4.2. Pilot study

A 1st year OSCE examination was obtained from the National University Ireland Galway (NUIG) School of Medicine. An internet search was conducted to identify freely available OSCE guides which specifically prepare medical students for their OSCE examinations to compare to the NUIG examination. The two most comprehensive guides found were from OSCE Skills and Geeky Medics [135, 136]. OSCE Skills and Geeky Medics were found mentioned across various online medical school forums, so we inferred that they are widely utilised by medical students internationally [137]. The guides were composed of a number of stations categorised into various medical specialties. Each station described the steps, actions, and considerations necessary for students to adequately perform the clinical scenario. The guides and examination were uploaded into NVivo 10.2.2. (QRS International, Melbourne, Australia), and each station was qualitatively analysed for the presence of SA [138].

A self-developed tool was used for this analysis. The tool was developed using Endsley's model of SA, it is constructed using specific tasks/goals assigned to each of the three levels of SA (Figure 4.2). The tool is an inferential measure of SA, meaning presence of SA is determined based on performance in the clinical scenario; if a station guide listed a task from the tool, the guide was said to possess that level of SA. Where there were multiple instances of a particular task in one station, it was recorded only once. Further, a single observation of SA in the station was recorded as one observation of SA in the tool, regardless of the length of text coded in NVivo. Multiple observers analysed the guides and examination using this tool. The qualitative results were then quantitatively analysed using Microsoft Excel. For descriptive statistics, we used frequency and measures of central tendency. For inferential statistics, we used Kruskal–Wallis for variance analysis, Cohen's kappa for interrater reliability, and Cronbach's alpha for internal validity.

4.3. Training of research participants

An introduction session was provided initially to research fellows (1 primary investigator, Senior Lecturer in Medical Informatics and Medical Education, 1 medical practitioner and 1 allied health care practitioner) to obtain a consistent comprehension of the model of SA by Endsley and characteristics attributed to SA in the medical context (Table 4.1). Subsequently, two freely available OSCE guides potentially preparing students for their evaluation of clinical skills (osceskills.com and geekymedics.com) were analysed and discussed openly utilising the self-developed node tree (Figure 4.2). Elements of the clinical assessment had to be appraised for the level of SA, and subsequently assigned to the most appropriate facet within this level. In case of identification of diverging elements of SA within one phrase, the text passage had to be split and coded individually to the selected facet. Headlines, description of images, expected learning outcomes or educational instructions within the OSCE forms were excluded form coding.

This was followed by the evaluation of further two freely available OSCE guides as an individual home-based exercise. Upon appraisal of the interrater disagreement identified in the home-based exercise, the researchers openly discussed any discrepancy to achieve optimal understanding of the meaning of parental nodes and child nodes. This was followed by successive evaluations of further four home-based exercises including two randomly selected freely available OSCE guides and OSCE score sheets utilised at the National University Ireland Galway (NUIG) respectively at any one time. All 8 forms were independent from the actual study. Progress in the level of interrater agreement was determined by calculating Cohen`s Kappa.

Ongoing disagreement was openly discussed to achieve concordance and final decisions were adjudicated by the principle investigator if deemed necessary. Upon proven increment in agreement between researchers three randomly selected freely available OSCE guides and three randomly selected OSCE forms of the medical training at NUIG (2015-2016) were analysed individually for the final study (Table 4.2). The random selection of OSCE guides and score sheets was obtained by utilization of the INDEX function within EXCEL with help from the RANDBETWEEN and ROWS functions. This method utilises the formula = RAND() to generate a numeric value between 0 and 1 for each recorded parameter. Random values were prevented from changing when the worksheet recalculates by selecting and copying all values in the column. Upon allocation of a random number to each row, the records were sorted by the Random_number column.

Table 4.2: Structure of the Consecutive Guided Training

Type of Training	Duration	Learning activity	Homework	Overall Level of Agreement (%)
Initial Introduction	120 minutes	Initial introduction to Endsley`s model of SA in high-risk environments and SA in the medical context		**N/A**
Introduction	120 minutes	Theoretical and practical introduction to NVIVO, open discussion of elements of SA (Fig. 2) in medicine identified in two freely available OSCE guides	Coding of two freely available OSCE guides (OSCEskills; Abdominal Examination, Greekemedics: Ankle and Foot Examination) by manual marking	**N/A**
Training Seminar	120 minutes	Identification of discrepancy in coding elements of SA in the homework exercise and additional instruction of interpretation of elements of SA in assessment forms	Coding of two freely available OSCE guides utilising NVIVO Greekemedics: Blood Pressure Measurement/ CXR Interpretation	Interrater A. **SA 1: 0.63/ 0.55** **SA 2: 0.62/ 0.72** **SA 3: 0.67/ 0.65**
Training Seminar	120 minutes	Review of two freely available OSCE guides and discussion and clarification of agreement and disagreement of the coding results	Coding of two OSCE forms previously used at NUIG utilising NVIVO NUIG- SOB NUIG- Anaphylaxis	Interrater A. **SA 1: 0.66/ 0.85** **SA 2: 0.61/ 0.69** **SA 3: 0.61/ 0.63**
Training Seminar	120 minutes	Review of two OSCE forms previously used at NUIG and discussion and clarification of agreement and disagreement of the coding results	Re-Coding of two previous coded OSCE guides utilising NVIVO Greekemedics: Blood Pressure Measurement/ CXR Interpretation	Intrarater A. **SA 1: 0.72/ 0.85/** **SA 2: 0.67/ 0.69** **SA 3: 0.71/ 0.63**
Training Seminar	120 minutes	Review of the re-coding of two previously coded OSCE guides and discussion and clarification of agreement and disagreement of the results	Coding of two freely available OSCE guides utilising NVIVO OSCESkills: CVS Examination/ Respiratory Examination	Interrater **SA 1: 0.69/ 0.73** **SA 2: 0.60/ 0.78** **SA 3: 0.55/ 1.00**
Concluding Training Seminar	120 minutes	Review of the two freely available OSCE guides and final discussion and clarification of agreement and disagreement in the coding results in preparation for the evaluation of six randomly selected OSCE guides/ forms of the interrater agreement for the study		**N/A**

4.4. Final test

Four independent researchers (1 primary investigator, Senior Lecturer in Medical Informatics and Medical Education, 1 medical practitioner and 1 allied health care practitioner) consecutively examined 6 randomised OSCE guides/ score sheets in a mixed (qualitative and quantitative) method. This mixed method utilised a self-developed node tree using NVIVO 10 software, allowing for coding of information to predefined nodes by multiple researchers and subsequent interrogation of diverse types of queries and comparisons (NVIVO version 10.0.638.0 SP6 32-bit) [138].

4.5. Data processing

Interrater agreement (Cohens Kappa) was calculated by NVIVO based on levels of agreement (%) and disagreement (%). The kappa coefficient is determined by the amount of the total units of agreement minus an expected frequency of random or accidental agreement, divided by the sum of total units for evaluation minus the expected frequency of random or accidental agreement. Interrater agreement was considered as very good if Cohen’s Kappa is > 0.80, good when ranging between 0.60 – 0.80 and moderate when ranging between 0.40 – 0.59. 1 – Cohen’s Kappa is the level of disagreement (error) between raters.

Generalisability theory (GT) analysis is additional to classical psychometric analysis and does provide insight in the different sources of error (around the observed score of agreement) [139]. GT consists of a Generalisability-Study (G-Study) and a Decision-Study (D-Study). In the G-Study the main facets of variation and all their interactions are being examined. The D-Study allows to calculate the effect of experimental measurement designs on the reduction of the error around the observed score [140]. The classical psychometric analysis does not provide insight into the causes of this level of disagreement as the percentage of agreement is not corrected for change and therefore, this correction is random.

The GT analysis supports the identification of the variability of sources of error around the observed score of agreement [141]. GT consists of a Generalisability-Study (G-Study) and a Decision-Study (D-Study). In the G-Study the main facets of variation and all their interactions are being examined. The D-Study allows to calculate the effect of experimental measurement designs on the reduction of the error around the observed score [140].

The Standard Error of Measurement (SEM) is described as the error around the observed score and is expressed in the same unit of measurement as the assessment tool used (% agreement). An observed score is the result of the unknown true score and error around the observed score. The true score is an optimal score out of the universe of potential scores [142].

To analyse the impact of the main facets on the variation between Levels of SA (L), the Items (I) embedded within these levels and the 4 Raters (R), we utilised EduG for the G-study and G-study analysis [143]. Generalisability Coefficient (G-coefficient) is considered as the reliability coefficient addressing agreement between examiners. The main facets of analysis were defined by the Analysis of Variance (ANOVA) as OSCE Forms (F), Individual Raters (R) and the 3 Levels of SA (L). Two consecutive designs of measurement were chosen to 1. Analyse the raters as object of measurement (R/FLI) and 2. Analyse the OSCE scoring sheets as object of measurement (F/RLI). Four items were associated with elements of SA, the fifth item is representing the total of that level of SA.

The Generalisability Coefficient (G-coefficient) is considered as the reliability coefficient addressing agreement between examiners [141]. A G-coefficient > then 0.8 was considered to be high; between 0.6 – 0.8 as moderate and below 0.6 as low agreement. The main facets of analysis were defined and analysed in an ANOVA (Analysis of Variance) as OSCE forms (F); individual Raters (R); and the 3 levels of SA (L). The items embedded within the levels of SA will be characterised as I:L ((I:L), items embedded within levels). Two way interaction effects between Forms and Raters (FxR); Forms and Levels (FxL); Forms and Items embedded in Levels (FxI:L) further two interaction between Raters x Levels (RxL); Raters interpreting Items within Levels (RI:L) have been analysed.

Three way interaction effects will be analysed as forms rated by different raters assessing different levels of SA (FxRxL) and finally the residual error Forms, Raters and Items within Levels (FRI:L). Two consecutive designs of measurement have been utilised for 1. Analysis of the OSCE scoring sheets as object of measurement (F/RLI) and 2. Analysis of the Raters as object of measurement (R/FLI). The Standard Error of Measurement (around the agreement score) is described as the error around the observed score and is expressed in the same unit of measurement as the assessment tool used (% agreement) [142].

Chapter 5: Results

5.1. Pilot study

Outcomes of the pilot study utilising this self-developed tool demonstrated that researchers were able to identify elements of SA, however, interrater agreement was only moderate. Upon an initial introduction to the model of SA and its adaption to healthcare, raters were able to identify elements of SA, however, only moderate interrater reliability has been demonstrated. The OSCE Skills guide and Geeky Medics guide consisted of 33 stations in nine medical specialties and 39 stations in ten medical specialties, respectively. The NUIG examination was significantly less comprehensive, consisting of only five stations, which were not categorised into any specific specialties. The specialties included in the OSCE Skills guide and the Geeky Medics guide were cardiology, endocrinology, gastroenterology, neurology, obstetrics and gynaecology, orthopaedics, otorhinolaryngology, paediatrics, psychology, pulmonology, urology, and an "other" category.

The NUIG OSCE examination possessed both absolutely and relatively fewer observations of SA as compared to the two free OSCE guides. On an average, 45% of the stations exhibited Level 1 SA, 18.9% of the stations exhibited Level 2 SA, and 21.7% of the stations exhibited Level 3 SA, with a standard deviation of 5%, 3.8%, and 12.6%, respectively. Figure 2 illustrates the comparison between the mean number of stations that demonstrated each level of SA within each OSCE guide and examination. The Geeky Medics guide exhibited the highest degree of SA in every level. Nearly 56.5% of the stations exhibited Level 1 SA, 37.5% of the stations exhibited Level 2 SA, and 34.3% of the stations exhibited Level 3 SA, with a standard deviation of 5.6%, 4.7%, and 13.4%, respectively. OSCE Skills guide was in the middle level comparing Level 1 and 2 SA, but it had the lowest mean observations of Level 3 SA. Around 54.5% of the stations exhibited Level 1 SA, 24.2% of the stations exhibited Level 2 SA, and 6.2% of the stations exhibited Level 3 SA, with a standard deviation of 12.1%, 11.3%, and 2.5%, respectively.

Table 5.1: Kruskall Wallis test results comparing variance between each rater for each guide at each level

Guide	Level 1	Level 2	Level 3
OSCE Skills	H=0.206(1, N=20), p>0.05, fail to reject Ho	H=0.516(1, N=20), p>0.05, fail to reject Ho	H=0.1.851(1, N=20), p>0.05, fail to reject Ho
Geeky Medics	H=0.439(1, N=18), p>0.05, fail to reject Ho	H=0.329(1, N=18), p>0.05, fail to reject Ho	H=4.68(1, N=18), p>0.05, reject Ho
Ho=not statistically different Ha=statistically different			

Table 5.2: Cronbach's alpha scores for internal validity of each guide at each level

Guide	Level 1	Level 2	Level 3
OSCE Skills	α =0.719	α =0.630	α =0.847
Geeky Medics	α =0.851	α =0.991	α =0.875

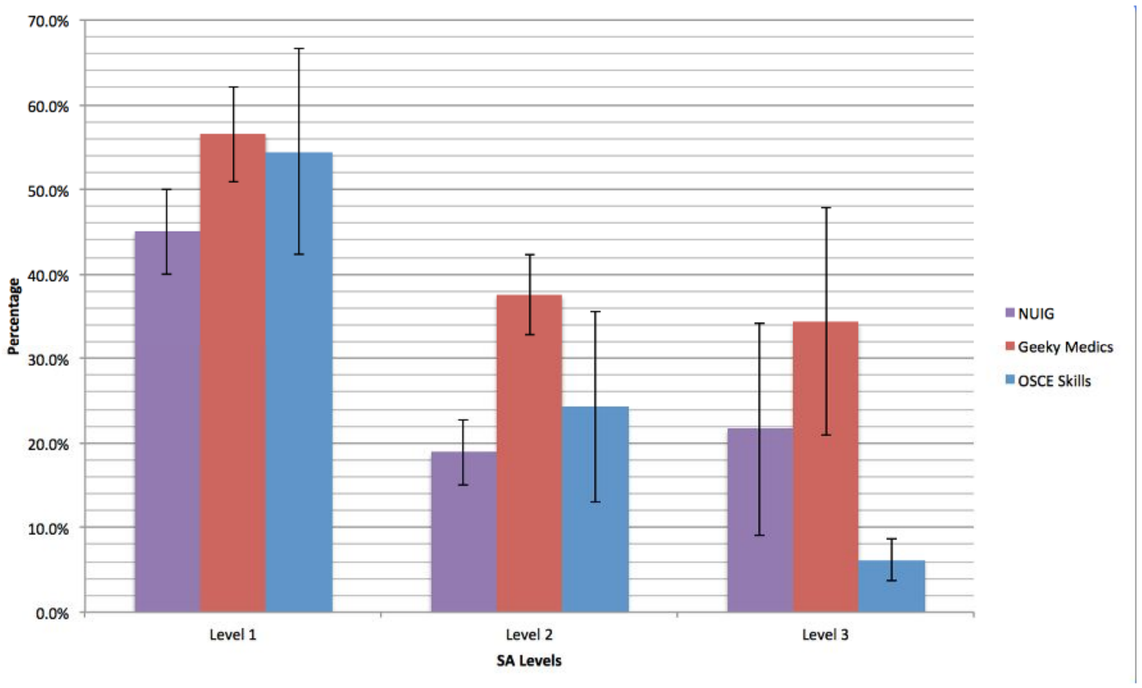


Figure 5.1: Mean frequency of SA observations in both guides and NUIG exam, with error bars representing standard deviation

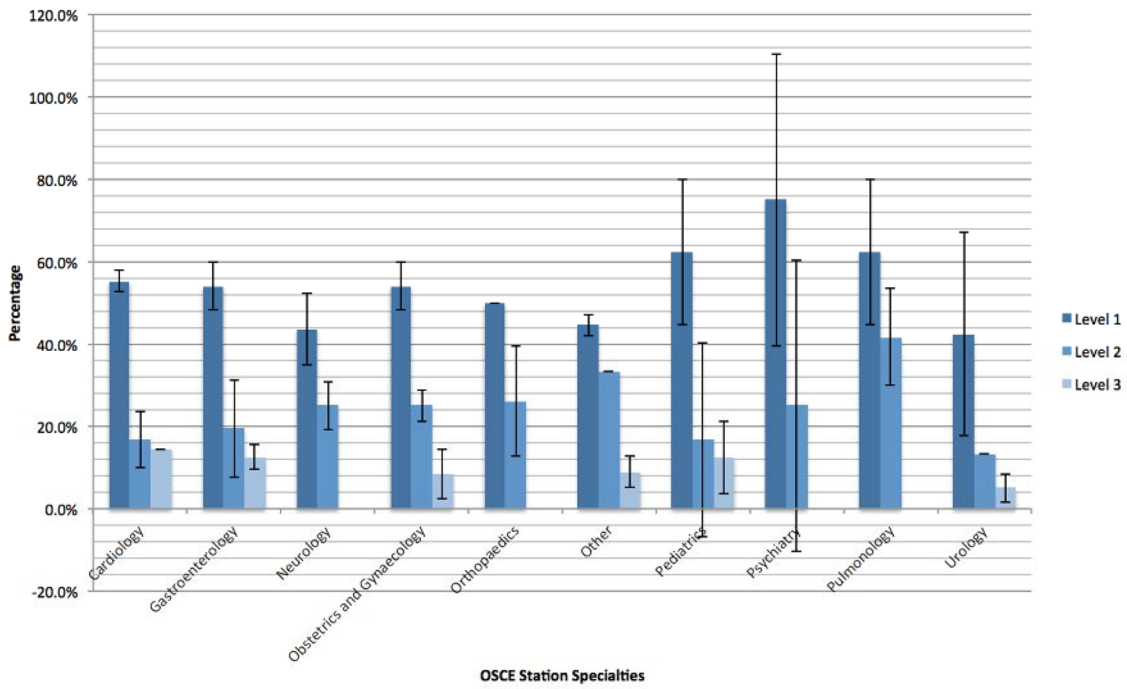


Figure 5.2: Mean frequency of SA observations within each specialty in OSCE Skills guide, with error bars representing standard deviation

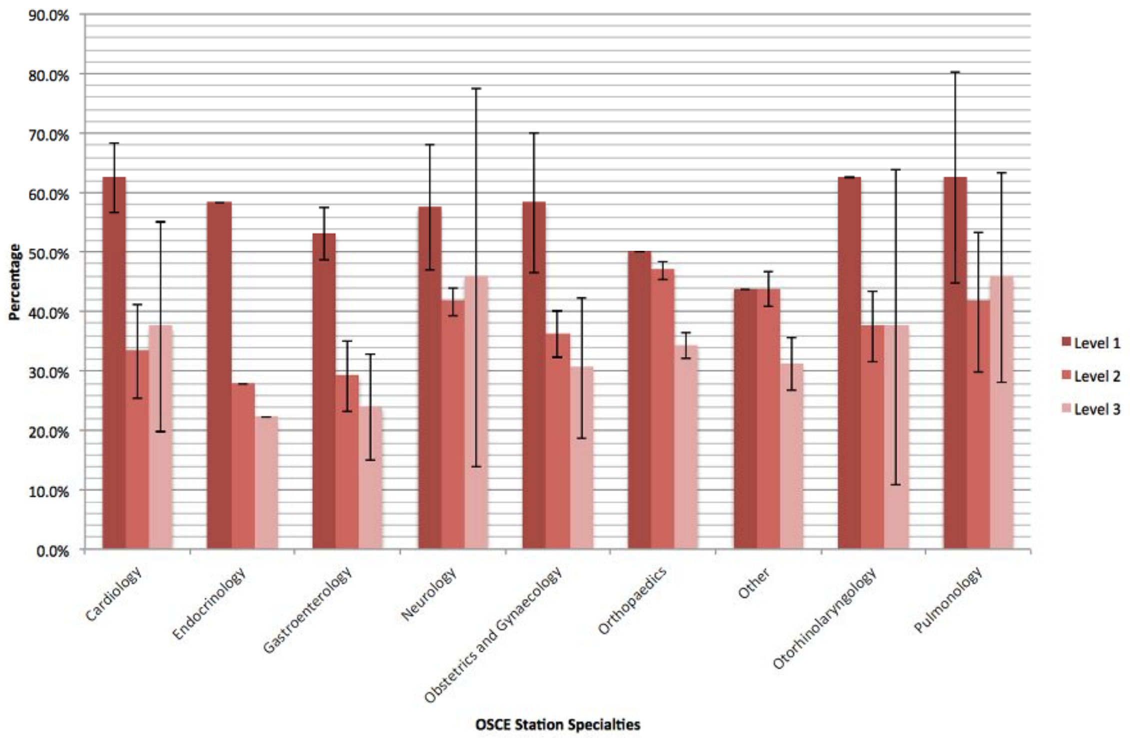


Figure 5.3: Mean frequency of SA observations within each specialty in Geeky Medics guide, with error bars representing standard deviation

5.2. Evaluation of consecutive training

Identifying and categorising elements of SA within OSCE forms by 4 individual researchers demonstrated a moderate to very good interrater agreement based on Cohens` Kappa (0.497 – 1.00) (Figure 5.4). The G-Theory revealed four key facets for variance: OSCE Forms/Scoresheet (F) (n=6); the Independent Raters (R) (n=4); the Levels of SA (L) (n=3) and the Items embedded in these Levels of SA (I:L) (n=5) (Table 5.3). The absolute G-coefficient of the reliability study was 0.92 as compared to the results of the classical psychometric analysis.

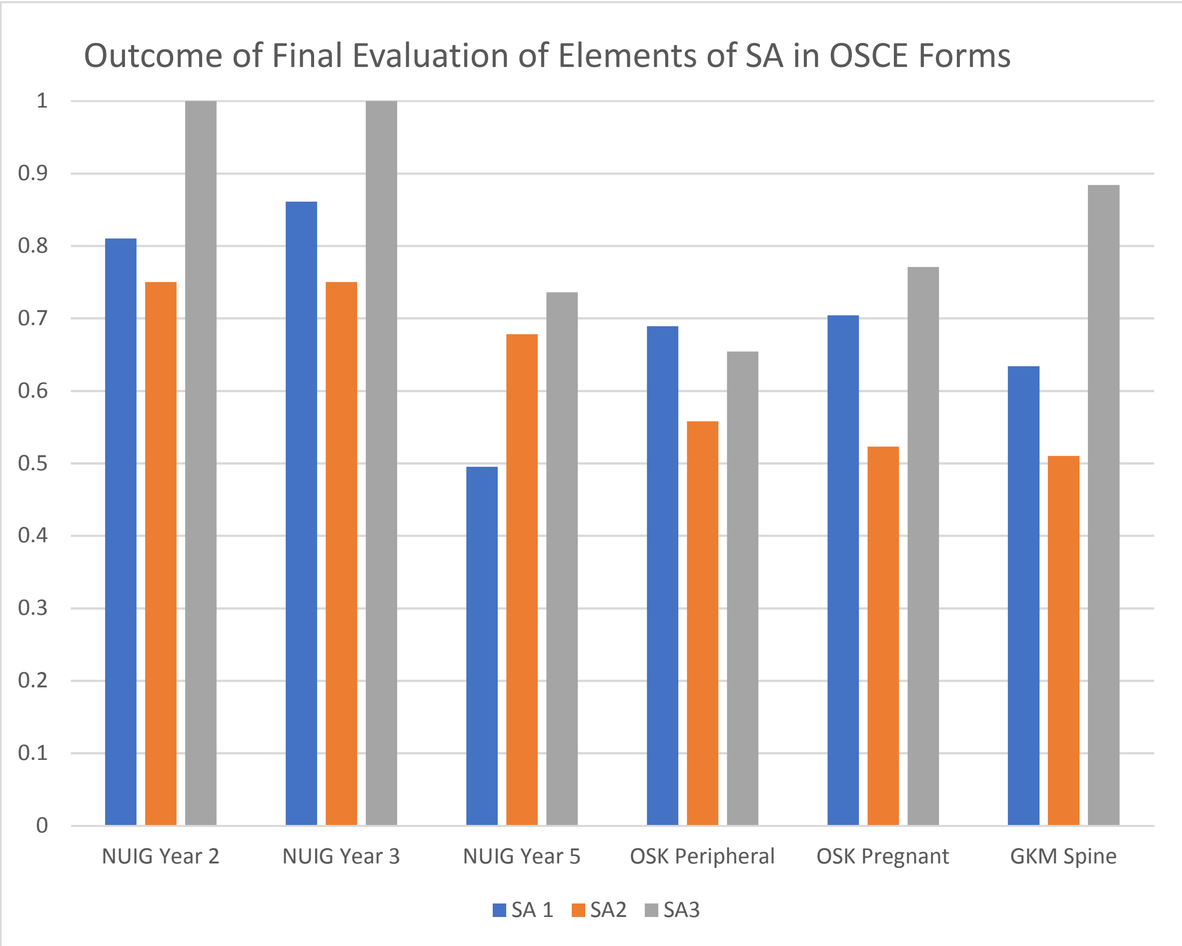


Figure 5.4: Overall interrater agreement (classical psychometric analysis) of levels of SA within six randomly selected OSCE forms between 4 independent raters

5.3. Results of the GT- Study

Of all variance, 2.7% is due to the OSCE score sheets, 0.4% is due to individual raters. 8.9% of variance can be attributed to the distinct levels of SA. Most of the main facets responsible for the variance were associated with the 'Items embedded in the Levels' seen as high as 32.7%. Furthermore, 0.6% of variance is due to the effect of the interaction between Forms and Raters (raters being influenced by the different types of forms), 15.2% are associated with interaction between Forms and Levels and 20.3% with Forms and Items embedded within Levels. Additionally, small interaction effects were identified with a residual unexplained error of 9.3%.

Table 5.3: Facets and associated labels and levels for the study design

Facet	Label	Levels	Univ.
Forms	F	6	INF
Raters	R	4	INF
Levels	L	3	3
Items in levels	I:L	5	5

Table 5.4: Analysis of variance incorporating all facets

(SS = sum of squares, df = degree of freedom, MS mean square, SE = standard error)

Analysis of variance

				Components				
Source	SS	df	MS	Random	Mixed	Corrected	%	SE
F	125.02222	5	25.00444	-0.35422	0.39407	0.39407	2.7	0.40062
R	18.26667	3	6.08889	-0.06252	0.05254	0.05254	0.4	0.09116
L	589.08889	2	294.54444	0.75707	1.96104	1.30736	8.9	1.81590
I:L	1917.95000	12	159.82917	6.01981	6.01981	4.81585	32.7	2.51926
FR	20.40000	15	1.36000	-0.17378	0.09067	0.09067	0.6	0.07306
FL	488.64444	10	48.86444	1.64765	2.24489	2.24489	15.2	1.00584
FI:L	799.25000	60	13.32083	2.98620	2.98620	2.98620	20.3	0.59921
RL	85.93333	6	14.32222	0.27743	0.34519	0.34519	2.3	0.24244
RI:L	122.71667	36	3.40880	0.33880	0.33880	0.33880	2.3	0.13254
FRL	119.00000	30	3.96667	0.51813	0.79333	0.79333	5.4	0.20042
FRI:L	247.68333	180	1.37602	1.37602	1.37602	1.37602	9.3	0.14425
Total	4533.95556	359					100%	

5.3.1. Raters as object of measurement

Using a measurement design in which the main sources of variation are the raters it appeared that 81.3% of the variation is due to the forms whereas 18.7% is due to the raters with an overall G-coefficient of 0.39. An assumptive increase in the amount of forms being analysed suggested an increment of the reliability with the G-coefficient raising from 0.39 to 0.65. The SEM can experimentally be reduced from 28% to 14% when quadrupling the number of OSCE forms to be analysed. The G-facets analysis based on raters as object of measurement demonstrated the level of unreliability for each individual OSCE form. In this measurement setting, the absolute G-coefficient for the 6 individual forms utilised in the final assessment ranges from 0.208 to 0.487, indicating low reliability of the OSCE forms. The G-coefficient for the individual levels SA 1, SA 2, SA 3 was calculated as 0.296, 0.032 and 0.000 respectively, indicating a poor reliability of these facets.

Table 5.5: Results of the applied G-Study based on raters (R) as object of measurement

G Study Table

(Measurement design R/FLI)

Source of variance	Differ-entiation variance	Source of variance	Absolute error variance	% absolute
	.....	F	0.06568	81.3
R	0.05254		.....	
	.....	L	(0.00000)	0.0
	.....	I:L	(0.00000)	0.0
	.....	FR	0.01511	18.7
	.....	FL	(0.00000)	0.0
	.....	FI:L	(0.00000)	0.0
	.....	RL	(0.00000)	0.0
	.....	RI:L	(0.00000)	0.0
	.....	FRL	(0.00000)	0.0
	.....	FRI:L	(0.00000)	0.0
Sum of variances	0.05254		0.08079	100%
Standard deviation	0.22922		Absolute SE:	0.28424
Coef_G absolute 0.39				

Table 5.6: G-Facets analysis based on Raters as object of measurement

G-Facets analysis

Facet	Level	Coef_G abs.
F	1	0.37
	2	0.47
	3	0.21
	4	0.34
	5	0.31
	6	0.49
L	1	0.30
	2	0.03
	3	0.00

Table 5.7: Results of the applied D-Study based on Raters as object of measurement

Optimisation

	G-study		Option 1		Option 2		Option 3	
	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.
F	6	INF	12	INF	18	INF	24	INF
R	4	INF	4	4	4	4	4	4
L	3	3	3	3	3	3	3	3
I:L	5	5	5	5	5	5	5	5
Coef_G abs.		0.39407		0.48239		0.58298		0.65083
rounded		0.39		0.48		0.58		0.65
Abs. Err. Var.		0.08079		0.04228		0.02819		0.02114
Abs. Std. Err. of M.		0.28424		0.20563		0.16790		0.14540

5.3.2. OSCE forms as object of measurements

The measurement design analysing the impact of the OSCE forms (F/RLI) revealed that 36.7% of variance was related to the raters and 63.3% of variance was assigned to the forms being analysed. The optimisation using a D-study revealed that an increase in the number of raters (from 4 to 6, 8 and 10) analysing a fixed number of forms (6) only contributes to an increase of about 4% in reliability. The associated Standard Error of Measurement would improve from 19% (0.189) to 12% (0.119). The G-coefficient, indicating the reliability of each of the 4 raters varies between 0.84 to 0.96, suggesting an overall high reliability. The results for the individual levels of SA demonstrate Level 1 to be very reliable, Level 2 as not reliable and Level 3 as less reliable

Table 5.8: Results of the applied G-Study based on OSCE forms (F) as object of measurement

G Study Table
(Measurement design F/RLI)

Source of variance	Differ-entiation variance	Source of variance	Absolute error variance	% absolute
F	0.39407		.....	
	.....	R	0.01314	36.7
	.....	L	(0.00000)	0.0
	.....	I:L	(0.00000)	0.0
	.....	FR	0.02267	63.3
	.....	FL	(0.00000)	0.0
	.....	FI:L	(0.00000)	0.0
	.....	RL	(0.00000)	0.0
	.....	RI:L	(0.00000)	0.0
	.....	FRL	(0.00000)	0.0
	.....	FRI:L	(0.00000)	0.0
Sum of variances	0.39407		0.03580	100%
Standard deviation	0.62775		Absolute SE:	0.18922
Coef_G absolute 0.92				

Table 5.9: G-Facets analysis based on OSCE forms (F) as object of measurement

G-Facets analysis

Facet	Level	Coef_G abs.
Raters	1	0.87
	2	0.88
	3	0.95
	4	0.84
Levels of SA	1	0.83
	2	0.00
	3	0.33

Table 5.10: Results of the applied D-Study based on OSCE forms (F) as object of measurement

Optimization

	G-study		Option 1		Option 2		Option 3	
	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.
Forms	6	INF	6	INF	6	INF	6	INF
Raters	4	INF	6	INF	8	INF	10	INF
Levels (fixed)	3	3	3	3	3	3	3	3
Items in Levels (fixed)	5	5	5	5	5	5	5	5
Coef_G abs.		0.91671		0.94289		0.95655		0.96493
rounded		0.92		0.94		0.96		0.96
Abs. Err. Var.		0.03580		0.02387		0.01790		0.01432
Abs. Std. Err. of M.		0.18922		0.15449		0.13380		0.11967

Chapter 6: Conclusion and limitation

6.1. Summary of work

A pilot study was initiated by the author of the thesis to identify the validity and interrater reliability of his self-developed SA assessment tool and to determine the degree of SA present in several OSCE forms utilised for clinical performance measures. Therefore, both freely available OSCE guides preparing students for this type of assessment as well as OSCE score sheets from a single medical education institution were examined. Upon a one-off introduction to Endsley`s model of SA and its association to clinical practice, researchers were able to identify performance markers and task requirements which can be assigned to the individual levels of SA. This study used a self-developed tool to assess SA in several OSCE guides and indeed a statistical difference between the guides and their degree of SA was found. It is difficult to determine which guides were "best."

We considered the guide that had the most observations of SA at the highest level to be more successful in applying SA to their clinical scenarios. In this respect, the NUIG OSCE appears to be inferior to the freely available OSCE guides on the internet. However, we do not find this to be an accurate representation of the teaching institution and its performance, as higher year examinations could not be accessed at the time of study commencement. Geeky Medics was the superior OSCE guide, it possessed the most observations of SA at every level and therefore seems to prepare students in the best manner.

We believed the best OSCE guide was the one that had the greatest frequency of observations across all SA levels because each level requires a different degree of mental processing. When information is perceived, it is stored in working memory instead of relevant long-term memory stores, or other similarly relevant mechanisms [144]. This is Level 1 SA and it involves responding to the input of relevant data. Comparing this to Level 2 SA requires new information to be taken in as working memory and then combined with the existing knowledge, resulting in the recognition of significant data and the generation of a holistic picture of the situation in one's mind. One step above that is Level 3 SA, which requires taking the composite picture of the new information in one's mind and using higher processing centers to generate an accurate plan in a timely fashion [144].

Some consider the three SA levels hierarchal, meaning that Level 1 SA is required to develop Level 2 SA and Level 2 SA is required to develop Level 3 SA. This is not necessarily the case, Endsley stated that SA can be a linear, bottom‑up process but it can also be a top‑down goal‑driven process, which will be explored later. In our view, it was not surprising to find that the most frequent tier of SA observed was Level 1 and the least frequent level of SA observed was Level 3, because Level 1 SA requires only reactions to working memory inputs and is the least mentally taxing. Notwithstanding, this observation could also be explained by the potential limitations of the tool in assessing higher level SA, or the subjectivity in the raters.

It was intriguing to see certain specialties with a greater degree of SA than others. SA is often studied in anaesthesiology because it is a dynamic medical specialty in which substantial and rapid changes occur [145, 146]. The specialties that exhibited the highest degree of SA in this study were neurology, otorhinolaryngology, paediatrics, and pulmonology. This was somewhat predictable as these specialties are equally dynamic—perhaps more so in paediatrics and otorhinolaryngology as both involve surgical care. Operating rooms are complex environments and have numerous people communicating, and different tools being used at the same time, thus requiring a high degree of SA [146]. While all medical specialties require SA, it is perhaps of increased importance in the above‑listed specialties, as reflected in the results of this study.

Our assessment tool demonstrated strong internal validity but moderate inter-rater reliability. This suggests that SA can indeed be assessed in OSCE guides, but there may be inherent subjectivity of the tool. It is possible to improve inter-rater reliability and increase objectivity of the tool with training sessions to optimise increase objectivity of the tool with training sessions to optimise use of the tool. Another limitation of the tool is that it is an indirect measure of SA.

Indirect measures infer SA, whereas direct measures are employed during simulation and are perhaps better indications of SA because they can explore an individual's thought process through in-test probes [147]. This method could more accurately predict SA performance, especially higher levels of SA. However, using a direct measure was beyond the scope of this study as we looked only at the guides and marking schemes for simulations, not the implementation of simulations. Ultimately, there is inherent difficulty in measuring SA, one study attempted to compare measures of SA using reliability and validity testing and it showed limited correlation, similar to our own study [147].

While the outcome of the study indicated a strong internal validity of the assessment tool, only a moderate level of interrater agreement has been identified. Therefore, this initial research suggested that further training would be essential to obtain a comprehensive understanding of the concept of Endsley in assessments of medical students. Furthermore, additional testing of the validity and reliability needs to demonstrate to what extent this self-developed tool can accurately measure or infer SA in OSCEs.

Following up on the results of the pilot study, the author developed a study protocol for a consecutive guided training programme to investigate the impact on subsequent validity and reliability measures. Therefore, co-researchers were introduced to the concept of SA and its applicability in medical practice. Upon a collective evaluation of freely available OSCE guides demonstrating and discussing how elements of SA can be identified using the self-developed assessment tool, coding of randomly selected OSCE guides and OSCE score sheets was carried out as individual home-based exercises.

The interrater agreement (Cohens Kappa), expressed by the levels of agreement (%) and disagreement (%) [138] was used as improvement measure for the beneficial impact of the training seminars. However, this classical psychometric analysis does not provide insight into the variance of disagreement. The need for an underlying understanding of facets potentially contributing to disagreement between raters seen in the pilot study suggested the application of the Generalisability Theory. This type of analysis incorporates both the Generalisability-Study (G-Study) and the Decision-Study (D-Study) [141, 142]. In the G-Study the main facets of variation and all their interactions are being examined. The D-Study allows calculation of the effect of experimental measurement designs on the reduction of the error around the observed score [140].

The outcome of the training was subsequently evaluated and findings were critically appraised for their ability to demonstrate improvement in interrater agreement as well as to indicate facets contributing to disagreement.

To the best of our knowledge, this was the first study to evaluate the impact of a consecutive guided training on how different raters are able to identify levels of SA. The outcome suggests that providing training enables educators and examiners to understand the concept of SA and to identify elements of SA within medical performance and competency assessments. That goes for each level of SA based on Endsley`s model being perception of situational elements, comprehension of elements in situation and the projection of their meaning for future development of that situation [148]. All OSCE score sheets used in our samples were designed without incorporating any specific knowledge or training in SA. We picked a random selection of forms of freely available OSCE guides and OSCE score sheets from a single medical curriculum of which no evidence showed that SA was part of the curriculum e.g. part of the assessment. However, the results of our study revealed that the OSCE Forms and the Items embedded in the individual levels of SA are not reliable for the purpose of assessment of SA.

The results for the individual levels of SA demonstrated Level 1 to be very reliable, Level 2 as not reliable and Level 3 as less reliable. The low occurrence of elements which can be attributed to the Levels 2 SA and 3 SA within the 6 OSCE forms utilised for the study might be causative for the poor outcome. Compared with the outcome of the preliminary study [149], Cohens-Kappa in our evaluation demonstrated an improved outcome of interrater agreement, ranging between moderate and very good levels of agreement. This suggests that the consecutive guided training provided to researchers had beneficial impact. The G-theory revealed no significant improvement of the results by the addition of further raters (G-coefficient raised from 0.92 to 0.96 when doubling the number of raters). In contrast, the addition of OSCE guides and score sheets did show an improvement of interrater agreement (from 0.39 to 0.65 when quadrupling the number of OSCE Forms). The amendment of the SA score description was identified as one key contributor to a superior outcome. A clear instructional outline of the expected activities and behaviour is suggested to support the intelligibility of OSCE score sheets by individual raters, thereby fostering the standardisation of the qualitative and quantitative evaluation of cognitive processing within clinical competence examinations.

6.2. Conclusion

Assessment of the development of clinical expertise remains a challenge to medical education as cognitive performance cannot be evaluated by direct observation [75]. Furthermore, fundamental cognitive processes in developing clinical expertise by medical students have not been clearly identified. This results in a lack of instructional measures enabling the development of the cognitive competence as part of CR in medical students [70]. The necessity of the ability to categorise assessment criteria of the OSCE score sheets into the elements amongst each level of SA attenuates the need for training of medical examiners. Rater-based assessments have been identified as possibly biased and interrater reliability as poor [150].

Raters are influenced by own cognitive and perceptual abilities and limitations when assessing testees which might impact the quality of their judgment of students' performance [151]. This highlights the need for assessors to be able to adequately identify cognitive abilities as one cornerstone of clinical competence. Based on research outcomes in underlying science of diagnostic errors, Singh et al. recommended the reconfiguration of training and education as well as the development of assessment methods to measure the quality of diagnostic care [152].

An analytical tool to identify breakdowns in SA in the underlying diagnostic process could differentiate elements within the clinical encounter which can be categorised into the level of SA [74]. Whole-task OSCEs are suggested to enable the evaluation of the utilisation of SA [104]. Fida and Kassab indicated that scores achieved by medical students in OSCE stations strongly correlated with the students´ ability to select and incorporate pertinent information and competence in patient management [8]. The summative evaluation of the integration of various competencies by individual assessors might facilitate a scoring system enabling the inferring of the underlying cognitive process of medical students [153]. For example, the satisfactory completion of a thorough history taking or physical examination by the student suggests an adequate Level 1 SA. Subsequent formulating of an incongruously working diagnosis, however, might suggest a flawed incorporation of the gathered information in subsequent cognitive processing correlating with deficiencies in Level 2 SA.

Wilkinson et al. demonstrated a correlation between direct involvement of the examiners in the designing of OSCEs and interrater reliability. Collaboration in the development of clinical assessment stations including objectives, format and score sheets were suggested to improve subsequent examiners understanding [154]. However, clinicians who developed their expertise over many years are commonly unaware of the levels of SA and, thus, they generally cannot convey or teach this process of data gathering and incorporation into the judgmental process [68]. Our study demonstrated that clinical practitioners and medical educators can be trained in understanding of the meaning of elements of SA in the medical context identified in assessment forms. Though, identification of key elements of cognitive competencies within medical assessments was demonstrated as being difficult.

6.3. Limitations of the Study

The first limitation of this study is seen in the minimal number of randomly selected OSCE forms, as one outcome indicated that the level of interrater agreement would improve when evaluating more documents. An additional limitation was the utilisation of the self-developed assessment tool by lack of an existing tool incorporating valid and reliable behavioural activities which can be attributed to the individual levels of SA. The breadth of the model is great, allowing application to multiple industries, including health care and the many fields within health care [155].

As previously mentioned, the model is not unidirectional, it can be understood as a forward mechanism in decision-making process or a backward mechanism in goal-driven processes [146]. Finally, new situations do not have to be the same as previous situations to employ SA. SA is a nontechnical skill developed over time and allows cues to be recognized, regardless of the circumstances [144]. The model represents a dynamic cycle of collecting, interpreting, and predicting information in any condition, which is why it was used to develop the tool for assessing SA in this study. Furthermore, raters were selected from different medical background and disciplines which does not allow for any conclusion for specific roles within the medical education faculty.

Chapter 7. Generalisation

7.1. Preface

SA has been identified as a vital skill for medical practice, at the same time no assessment concept to measure the level of SA validly has been presented. While training and assessment of SA are increasingly incorporated into medical practice, evidence on how to best train and teach students in obtaining and maintaining consciousness of a given situation in medical education is largely lacking. Our study shows that elements identified in OSCE score sheets can be assigned to the levels of SA.

Simulated patient encounters may help in the development of a cognitive map and thereby, give insights into information processing among medical students. Thus, deficits in recognising and incorporating essential parameters during the assessment can be identified and remediated when developing clinical expertise. This potentially can prevent the necessity of tackling habits already evolved over time. Further research is necessary to improve the assessment of SA and to determine to which degree OSCE assessment forms can be utilised to identify where the chain of SA was broken down. Ultimately, it may help in facilitating the development of educational strategies fostering cognitive reasoning abilities among medical students.

7.2. Aim of study

The purpose of the thesis was to explore the adaptability of Endsley`s model of SA to OSCE assessments in undergraduate medical education. This model is widely applied in teaching and assessing performance in high-risk environments, such as aviation, oil platforms and nuclear power plants [26]. Practical simulation exercises in these areas facilitate to impart the importance of obtaining the “big picture” for the safe completion of the given mission mirroring critical situations [37].

In healthcare, the accentuation of the significance of SA for effective task performance is slowly incorporated into clinical specialist training programmes [86, 117, 156]. Reports for the implementation of SA into undergraduate medical curricula are a fortiori [46]. The key focus of this research project was directed towards the potential usability of the OSCE as teaching and assessment method for SA at undergraduate level. Therefore, three studies investigated the feasibility and potential obstacles for the integration of measures for SA within these clinical performance assessments in undergraduate medical curricula.

The outcome of this thesis is to provide recommendations for developing educational means to enhance students` ability to collect, select and integrate relevant information. The potential to identify the fundamental cognitive processing of data could implement strategies to foster the ability to obtain and maintain SA during clinical encounters. The thesis provides answers regarding the availability of elements of clinical practice which can offer valuable insights whether levels of SA can be identified in OSCEs. The self-developed assessment tool for process- and outcome-based criteria for SA could advance into a valid and reliable method to identify students` cognitive competency in information processing. Therefore, three research questions were formulated as follows:

1. The purpose of the literature search is to review the literature with a view to identifying whether levels of SA can be assessed during undergraduate medical training utilising OSCEs based on Endsley's model

2. The purpose of the pilot study is first to determine validity and next to determine interrater reliability of a SA assessment tool and to determine the degree of SA present in several medical student OSCE guides.

3. The purpose of the study is 1. to evaluate the effect of consecutive 'guided' training on the improvement of interrater agreement, 2. to assess the reliability of a method for identifying elements of SA embedded in Objective Structured Clinical Examination (OSCE) station score sheets, which can be categorised to the three levels of SA (Level 1 SA, Level 2 SA, Level 3 SA) based on Endsley's model, and 3. to identify facets contributing to interrater disagreement.

7.3. General discussion

This thesis provides fundamental particulars of the existence of elements within clinical practice which can be assigned to the individual levels of SA. There is ongoing debate between scientists about whether SA represents either the underlying process itself or the outcome of the process when obtaining the “big picture” of what’s going on [25, 52]. Therefore, there is no generally accepted measurement instrument to evaluate SA [37, 50]. However, irrespective of the definition by researchers, the development and utilisation of SA depends on the individual cognitive information processing. To make matters worse, SA in medical practice has gained little importance and acceptance over the last decades.

Diagnostic errors and faulty treatment decisions were more likely attributed to overlying system errors rather than individual cognitive misconduct [11, 12]. However, increasing numbers of malpractice claims resulting in degrading reputations of clinicians and increasing expenses of healthcare providers due to erroneous clinical decisions underpin the necessity for medical faculties to develop methods to improve fundamental cognitive information processing [82].

The model of Endsley defines SA as an interdependent three-levelled concept, which is widely accepted in simulation research in critical environments [66]. The OSCE is commonly the only practical simulative assessment of students` clinical skills and performance integrated into the undergraduate medical curricula. This suggests the opportunity to compare the potential of the OSCE to evaluate students` development and utilisation of SA similarly to assessment methods used in high-risk environments.

7.3.1. Situation awareness in OSCE assessments

The OSCE provides an opportunity to assess characteristics of clinical performance which can be assigned to the individual levels of SA as adapted from the model by Endsley. Whole-task OSCEs are suggested to enable the evaluation of the utilisation of situational elements in CR in medical students in a staged format [101]. Typical performance markers such as the completion of physical examination and history-taking are commonly included in the OSCE score sheet and generally carried out successfully. However, students often lack the ability to sequentially embed their clinical findings into the proximate CR steps [72]. To overcome the mismatch between adequate information gathering and efficient data integration, development of educational measures to enhance cognitive processing are recommended.

The literature search for the first paper demonstrated that evaluation of Level 1 SA, as described by physical examinations, history taking but also in obtaining an overall impression of the patient and the retrieval of diagnostic test results, were found in all publications for the final assessment. Furthermore, all included studies demonstrated evaluation of elements of Level 2 SA, when incorporating the gathered information in subsequent information processing steps. However, assessment of activities which can be assigned to Level 3 SA were spare. Students` indicated that they appreciate this type of assessment as a means of valuable feedback for identifying areas of clinical weakness.

Accepting the importance of SA as an underlying requirement for well-informed CR inspired scholars to develop efficient information processing skills and techniques. Furthermore, it was highlighted that direct feedback from examiners provided upon completion of OSCE scenarios could support the faculty's appraisal and the examinees' self-rating of deficiencies within the sense-making process when selecting best clinical diagnosis and therapeutic options [124]. Though, providing individualised feedback based on performance was described as being complex [125]. In order to provide instructive recommendations based on the critical appraisal of the clinical performance, educators and assessors demand a fundamental background knowledge of the scoring methods and its associated informative value.

7.3.2. Situation awareness in OSCE guides and score sheets

Analysis of research outcomes in diagnostic reasoning errors and clinical errors direct the focus on the development of educational strategies which can be implemented into early medical training including assessments such as the OSCE [116]. This necessitates the ability to assign performance criteria of the utilised score sheets to the associated level of SA. Outcomes of a preliminary study, using a self-developed tool to assess SA in several OSCE guides and OSCE score sheets revealed a statistical difference in the quantitative identification of levels of SA between the various OSCE guides and OSCE score sheets [149]. OSCE guides with a high frequency of observations across all levels of SA indicated higher requirements for cognitive information processing.

Utilisation of the assessment tool enabled individual raters to identify elements of SA. However, the moderate interrater reliability might demote the tool due to its inherent subjectivity. To improve interrater reliability and thus, the objectivity of the assessment instrument, a guided training to optimise the comprehension of SA in clinical assessments of medical students was desirable. Upon attending a consecutive guided training, four co-researchers independently assessed and evaluated randomly selected OSCE forms. The outcome of that study suggested that clinical tutors and medical educators did benefit from additional training, indicated by an improved understanding of the meaning of elements of SA.

Further investigations of interrater disagreement, however, revealed factors contributing to difficulties in differentiating key elements of cognitive competencies. For example, a clearer characterisation of scoring parameters for performance and behaviour was suggested to result in an improved intelligibility of OSCE score sheets by individual raters. Thus, the pivotal facet was the amendment of the SA score description which would contribute to a superior outcome. A standardised scoring system could foster the qualitative and quantitative evaluation of cognitive processing within clinical competence examinations. Furthermore, an improved learning curve in raters was suggested by adding further OSCE guides and score sheets to the study design.

7.4. Generalisation

The aim of the course of studies in medical education research is to foster the development of an overall professional competency. Inadequate SA was identified as a primary parameter associated with deficient clinical performance, recommending the implementation of SA training including simulation at undergraduate level as realised in aviation [44, 45]. Endsley`s model is described as an ascending build-up of SA, however, without a direct linearity between the individual levels. The model facilitates the assessment of processes and products involved to obtain (Level 1 SA) and maintain (Level 2 SA and Level 3 SA) the "big picture". In addition, the construct enables to identify, how the product (state of knowledge and understanding of the situation) initiates further processes in order to improve the overall comprehension of the given circumstances.

Due to the missing experience and expertise, medical students are initially data-driven information gatherers. The novelty or uncertainty of the situation might overwhelm them with information, which needs to be evaluated for their significance in the given presentation of the patient's clinical condition and circumstances. Thus, this approach is not an effective mechanism. If more experienced, the goal-driven approach enables for the utilisation of the initial SA (comprehension and projection as described in Level 2 SA and Level 3 SA) to identify the need for a more focused search for expedient information. In CR, this mechanism supports the employment of reasoning strategies such as confirmation bias, heuristics or rule-out-worst-scenario. Once the initial level of SA is complete, maintaining SA necessitates an ongoing update based on additional or novel data. Combatting the utilisation of flawed heuristics or biased decision making in medical practice is generally a critical endeavour. Therefore, various situational elements must be embedded into the consistency of the cognitive process. This can be supported by early development of strategies for pattern scan or the route for efficient data acquisition. Thus, conveying the significance of SA at undergraduate levels might mitigate the genesis of these defective CR strategies in subsequent medical practice.

One method mirroring educational practice in high-risk environments can be seen in exposing medical students to clinical simulation scenarios such as the OSCE. This type of assessment was suggested to assimilate students` cognitive performance from the educational environment to clinical practice [157]. However, evaluation of progression in gathering information and subsequent comprehension of these data remains a challenge to medical education as fundamental cognitive processes in developing clinical expertise have not been clearly identified [70]. In addition, an overall accepted correlation between OSCE scores and student's clinical reasoning ability has not been identified [96]. This subsequently entails a lack of instructional measures to improve cognitive skills as part of CR in medical students. Due to its novelty in undergraduate medical curricula, literature does not reveal a method to evaluate students´ utilisation of SA in a simulated patient encounter.

While formative or summative OSCEs are suggested as a valuable means to teach SA [147, 158], the challenge remains how to best convey the significance of adequate SA for efficient task performance in healthcare education. Furthermore, the lack of a generally accepted definition and concept of SA results in the debate of the most appropriate assessment method. The summative evaluation of the integration of various competencies might facilitate a scoring system enabling the inferring of the underlying cognitive process of medical students. Whole-task OSCEs, having the potential for a comprehensive evaluation of students' clinical performance, could facilitate both, direct and indirect methods for measuring SA of students similar to simulation exercises in high-risk environments. It was highlighted that rating students by observational performance markers may only estimate behaviour-based SA [159]. However, Endsley`s model demonstrated practicability in measuring SA and associated validity and reliability as it incorporates the associated information processing steps.

The combination of direct and indirect measures for the development and achievement of SA can be applied. Thus, the OSCE enables for both the evaluation of processes and the product in SA. For example, the satisfactory completion of a thorough history taking or physical examination by the student suggests an adequate Level 1 SA. Subsequent formulating of an incongruously working diagnosis, however, might suggest a flawed incorporation of the gathered information in subsequent cognitive processing correlating with deficiencies in Level 2 SA. Planning subsequent steps for the treatment of the patient or additional interventions adds activities and products assigned to Level 3 SA.

However, comparing methods to measure SA applying reliability and validity testing indicated limited correlation [147]. Direct measures (i.e. SAGAT or SART) can be applied when stop and hold during an encounter exercise at prespecified milestones. This does provide the opportunity to identify whether the student gathered and incorporated all available information necessary for the comprehension of the situation. Furthermore, students` ability to embed the collected data into further cognitive processing as part of CR can be evaluated in a stepwise manner. Individual, task-directed or goal-directed feedback in between supports the student self-awareness when identifying lack of gathering and incorporating of relevant information. Employing the “think aloud” technique in the simulated patient encounter offers a possibility to perceive students’ utilisation of elements of SA. Furthermore, having the simulating patient trained as an observer to rate the behaviour or activities of the participants might enable the estimation of the obtained level of SA as suggested by the SPAM method.

All evaluation methods require that medical tutors and examiners do understand the model of SA and its application in clinical performance exercises. The ability to infer the level of SA from associated performance criteria of OSCE score sheets is a fundamental requirement. Direct measures are typically considered to be product-oriented as these techniques assess the outcome of SA. Indirect measures orientate the outcomes on the underlying process to achieve SA. Simulation scenarios have been suggested to inspire students to improve their cognitive processes as well as to enhance active learning which subsequently can be reflected in a reduced occurrence of medical errors [122, 158].

7.5. Implications for further research

If set up as a whole-task assessment, the OSCE demonstrated the feasibility to teach medical students the importance of SA in their daily clinical practice. The development of OSCEs in a staged format in way of a comprehensive patient encounter potentially allows for the incorporation of all three levels of SA. It has been highlighted that the positive learning experience for medical students avails from direct and focused feedback to obtain an immediate understanding of the concept of SA in a given scenario.

In order to convey the construct of SA based on Endsley`s model, educators, clinical tutors and examiners in undergraduate medical curricula must be able to identity elements of clinical practice which can be assigned to levels of SA. Furthermore, development and acceptance of methodologies to teach and assess students` utilisation of SA was demonstrated to improve by direct involvement of assessors and tutors. Development of score sheets for whole-task OSCEs should be guided by incorporation of the assessment of elements of SA associated with essential steps of clinical reasoning. As demonstrated by the G-Study and D-Study in our research, descriptions of performance markers in OSCE score sheets which can be assigned to levels of SA need to be more articulative. Thus, further research should aim to confirm the beneficial impact of a guided consecutive training seen in our study including a larger cohort of participants from different medical undergraduate faculties.

Inferring SA from performance marked within whole-task OSCEs offers the potential to identify gaps in the cognitive information process which subsequently could facilitate the provision of situated feedback to the student. At the same time, the potential to draw a map of students` cognitive information processing might enable faculties to obtain an understanding of how medical trainees deal with novel situations or uncertainty. Further research needs to identify how this could support the development of instructional teaching methods to convey fundamental concepts in obtaining and maintaining SA at an undergraduate level.

Subsequently, new developed OSCE score sheets must be evaluated for their validity and reliability in measuring the utilisation of SA inferred from performance markers. This can include the assessment of medical students` approach to problem solving when exposed to novel clinical presentations and situations with uncertainty.

As for the greater outlook, the adaptation of already developed electronic assessment tools evaluating students` performance in OSCEs to incorporate an immediate online statement of the underlying cognitive processing is envisaged.

7.6. References

1. La Pietra L, Calligaris L, Molendini L, Quattrin R, Brusaferro S. Medical errors and clinical risk management: state of the art. Acta Otorhinolaryngologica Italica. 2005. 25(6):339-346.
2. Kohn TL, Corrigan JM, Donaldson MS. To Err Is Human Building A Safer Health System. 2000. Washington DC: National Academic Press.
3. World Health Organisation; Patient Safety: Data and Statistics. Available from: http://wwweuro.who.int/en/health‑topics/Health‑systems/patient‑safety/data‑and statistics. (Cited on 2016 Aug 09).
4. Graber MR, Gordon R, Franklin N. Reducing diagnostic errors in medicine: what's the goal? Academic Medicine. 2002. 77(10):981-992.
5. Graber ML, Wachter RM, Cassel CK. Bringing diagnosis into the quality and safety equations. JAMA. 2012. 308(12):1211-1212.
6. Nendaz M, Perrier A. Diagnostic errors and flaws in clinical reasoning: mechanisms and prevention in practice. Swiss Med Wkly. 2012. 142(w13706):1-9.
7. De Jong T, Ferguson-Hessler MG. Types and qualities of knowledge. Educational Psychologist. 1996. 31(2):105-113.
8. Fida M, Kassab SE. Do medical students' scores using different assessment instruments predict their scores in clinical reasoning using a computer-based simulation? Adv Med Educ Pract. 2015. 6:135-141.
9. Zwaan L, Thijs A, Wagner C, van der Wal G, Timmermans DR. Relating faults in diagnostic reasoning with diagnostic errors and patient harm. Acad Med. 2012. 87(2):149-156.
10. Zwaan L, Thijs A, Wagner C, van der Wal G, Timmermans DM. Design of a study on suboptimal cognitive acts in the diagnostic process, the effect on patient outcomes and the influence of workload, fatigue and experience of physician. BMC Health Services Research. 2009. 9(1):65.
11. Henriksen K, Brady J. The pursuit of better diagnostic performance: a human factors perspective. BMJ Qual Saf. 2013. 22(Suppl 2):ii1-ii5.
12. Stark M, Fins J.J. The ethical imperative to think about thinking - diagnostics, metacognition, and medical professionalism. Camb Q Healthc Ethics. 2014. 23(4):386-396.

13. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. Academic Medicine. 2003. 78(8):775-780.
14. Singh H, Schiff GD, Graber ML, Onakpoya I, Thompson MJ. The global burden of diagnostic errors in primary care. BMJ Qual Saf. 2017. 26(6):484-494.
15. Flin R, Maran N, Basic concepts for crew resource management and non-technical skills. Best Pract Res Clin Anaesthesiol. 2015. 29(1):27-39.
16. Flin R, Yule S, McKenzie L, Paterson-Brown S, Maran N. Attitudes to teamwork and safety in the operating theatre. The Surgeon. 2006. 4(3):145-151.
17. Patey R, Flin R, Cuthbertson BH, MacDonald L, Mearns K, Cleland J, Williams D. Patient safety: helping medical students understand error in healthcare. Qual Saf Health Care. 2007. 16(4):256-259.
18. Wisborg T, Manser T. Assessment of non-technical skills in the operating room-one assessment tool per specialty? Acta Anaesthesiol Scand. 2014. 58(7):773-774.
19. Papa L, Langland-Orban B, Kallenborn C, Tepas JJ, Lottenberg L, Celso B, Durham R, Flint L. Assessing effectiveness of a mature trauma system: Association of trauma center presence with lower injury mortality rate. Journal of Trauma. 2006. 61(2):261-267.
20. Hales B, Terblanche M, Fowler R, Sibbald W. Development of medical checklists for improved quality of patient care. International Journal for Quality in Health Care. 2008. 20(1):22-30.
21. Croskerry P, Nimmo GR. Better clinical decision making and reducing diagnostic error. Journal of the Royal College of Physicians of Edinburgh. 2011. 41(2):155-162.
22. Pelaccia T, Tardif J, Triby E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. Med Educ Online. 2011. 16: p. 10.3402/meo.v16i0.5890.
23. Eva KW. What every teacher needs to know about clinical reasoning. Med Educ. 2005. 39(1):98-106.
24. Graber ML. Educational strategies to reduce diagnostic error: can you teach this stuff? Adv Health Sci Educ Theory Pract. 2009. 14 Suppl 1:63-69.
25. Cooper S, Porter J, Peach L. Measuring situation awareness in emergency settings: a systematic review of tools and outcomes. Open Access Emerg Med. 2014. 6:1-7.
26. Endsley MR. Situation awareness in aviation systems. In Garland DJ, Wise JA, Hopkin VD (Eds.). Human factors in transportation. Handbook of Aviation Human Factors, 1999:257-276. Mahawah, NJ, US: Laurence Erlbaum Assoiciates Publishers.

27. Sarter NB, Woods DD. Situation Awareness: A Critical But Ill-Defined Phenomenon. The International Journal of Aviation Psychology. 1991. 1(1):45-57.

28. Rousseau RTS, Breton R. Defining and Modeling Situation Awareness: Aritical Review in A Cognitive Approach To Situation Awareness: Theory and Application, T.S. Banbury Simon, Editor. 2004. New York: Ashgate Publishing.

29. Breton R, Rousseau R. Situation awareness: A review of the concept and its measurement. DREV TR-2001-220 2001. Defence Research Establishment Valcatier.

30. Lo JC, Sehic E, Meijer SA. Explicit or Implicit Situation Awareness? Situation Awareness Measurements of Train Traffic Controllers in a Monitoring Mode, in Engineering Psychology and Cognitive Ergonomics: 11th International Conference, EPCE 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014. Proceedings, D. Harris, Editor. 2014, Springer International Publishing: Cham. 511-521.

31. Soto D, Silvanto J. Reappraising the relationship between working memory and conscious awareness. Trends Cogn Sci. 2014. 18(10):520-5.

32. Kanki BG, Helmreich RL, Anca J. (Eds). Crew Resource Management. 2010, Waltham: Academic Press.

33. Yim HB, Seong PH. A Quantitative Team Situation Awareness Measurement Method Considering Technical and Nontechnical Skills of Teams. Nuclear Engineering and Technology. 2016. 48(1):144-152.

34. Endsley MR. Theoretical underpinnings of situation awareness: a critical review. In Endsley MR, Garland DJ. Situation awareness analysis and measurement, 2000. Mahawah, NJ, US: Laurence Erlbaum Assoiciates Publishers.

35. Goss JR. Teaching clinical reasoning to second-year medical students. Acad Med. 1996. 71(4):349-52; discussion 348.

36. Sohn YW, Doane SM. Memory processes of flight situation awareness: interactive roles of working memory capacity, long-term working memory, and expertise. Human Factors. 2004. 46(3):461-75.

37. Endsley MR, Jones W. A model of inter and intrateam situation awareness: Implications for design, training and measurement. In McNeese M, Salas E, Endsley MR (Eds.). New Trends in Cooperative Activities: Understanding system dynamics in complex environments, HFES, 2001. 46-67. Santa Monica

38. Graafland M, Schraagen JM, Boermeester MA, Bemelman WA, Schijven MP. Training situational awareness to reduce surgical errors in the operating room. Br J Surg. 2015. 102(1):16-23.
39. Graafland M, Schijven MP. Situational awareness: You won't see it unless you understand it. Nederlands Tijdschrift voor Geneeskunde. 2015. 159(30).
40. Fore AM, Sculli GL. A concept analysis of situational awareness in nursing. Journal of Advanced Nursing. 2013. 69(12):2613-2621.
41. Leonard MM, Kyriacos U. Student nurses' recognition of early signs of abnormal vital sign recordings. Nurse Education Today. 2015. 35(9):e11-e18.
42. Wassef ME, Terrill E, Yarzebski J, Flaherty H. The Significance of Situation Awareness in the Clinical Setting: Implications for Nursing Education. Austin J Nurs Health Care. 2014. 1(1):1005.
43. Parush A, Campbell C, Hunter A, Ma C, Calder L, Worthington J, Frank JR. Situational Awareness and Patient Safety: A Primer for Physicians. The Royal College of Physicians and Surgeons of Canada. 2011. Ottawa.
44. World Health Organsization. Human Factors in Patient Safety Review of Topics and Tools. Report for Methods and Measures Working Group of WHO Patient Safety 2009.
45. Walton M, Woodward H, Van Staalduinen S, Lemer C, Greaves F, Noble D , Ellis B , Donaldson L, Barraclough B. The WHO patient safety curriculum guide for medical schools. Qual Saf Health Care. 2010;19[6]:542-546.
46. Armitage G, Cracknell A, Forrest K, Sandars J. Twelve tips for implementing a patient safety curriculum in an undergraduate programme in medicine. Medical Teacher. 2011;33[7]:535-540.
47. Cornell P, Townsend-Gervis M., Vardaman JM, Yates L. Improving situation awareness and patient outcomes through interdisciplinary rounding and structured communication. J Nurs Adm. 2014. 44(3):164-169.
48. Starmer AJ, Spector ND, Srivastava R, Allen AD, Landrigan CP, Sectish TC. I-pass, a mnemonic to standardize verbal handoffs. Pediatrics, 2012. 129(2):201-204.
49. Klein G. The recognition-primed decision (RPD) model: Looking back, looking forward. Naturalistic Decision Making. 1997:285-292.
50. Salmon PM, Stanton NA, Walker GH, Jenkins D, Ladva D, Rafferty L, Younget M. Measuring Situation Awareness in complex systems: Comparison of measures study. International Journal of Industrial Ergonomics. 2009. 39(3):490-500.

51. Endsley M.R, Selcon SJ, Hardiman TD, Croft DG. A Comparative Analysis of Sagat and Sart for Evaluations of Situation Awareness. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 1998. 42(1):82-86.
52. Wright M, Taekman J, Endsley MR. Objective measures of situation awareness in a simulated medical environment. BMJ Quality and Safety. 2004. 13(Suppl 1):i65-i71.
53. Durso FT, Bleckley MK, Dattel AR. Does Situation Awareness Add to the Validity of Cognitive Tests? Human Factors. 2006. 48(4):721-733.
54. Shelton CL, Kinston R, Molyneux AJ, Ambroseet LJ. Real-time situation awareness assessment in critical illness management: adapting the situation present assessment method to clinical simulation. BMJ Quality and Safety. 2013. 22(2):163-167.
55. Patel VL, Yoskowitz NA, Arocha JF, Shortliffe EH. Cognitive and learning sciences in biomedical and health instructional design: A review with lessons for biomedical informatics education. Journal of Biomedical Informatics. 2009. 42(1):176-197.
56. Lubarsky S, Dory V, Audétat MC, Custers E, Charlin B. Using script theory to cultivate illness script formation and clinical reasoning in health professions education. Canadian Medical Education Journal. 2015. 6(2):e61-e70.
57. Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, Harris P, Glasgow NJ, Campbell C, Dath D, Harden RM, Iobst W, Long DM, Mungroo R, Richardson DL, Sherbino J, Silver I, Taber S, Talbot M, Harris KA. Competency-based medical education: theory to practice. Medical Teacher. 2010. 32(8):638-645.
58. Lurie SJ, Mooney CJ, Lyness JM. Measurement of the general competencies of the accreditation council for graduate medical education: a systematic review. Acad Med. 2009. 84(3):301-309.
59. Carraccio CL, Englander R. From Flexner to competencies: reflections on a decade and the journey ahead. Acad Med. 2013. 88(8):1067-1073.
60. Kane MT. The assessment of professional competence. Eval Health Prof. 1992. 15(2):163-182.
61. Epstein RM. Assessment in Medical Education. N Engl J Med. 2007. 356(4):387-396.
62. Launer J. Unconscious incompetence. Postgraduate Medical Journal. 2010. 86(1020): p. 628-628.
63. Cutrer WB, Sullivan WM, Fleming AE. Educational strategies for improving clinical reasoning. Current Problems in Pediatric and Adolescent Health Care. 2013. 43(9):248-257.

64. Pinnock R, Welch P. Learning clinical reasoning. Journal of Paediatrics and Child Health. 2014. 50(4):253-257.

65. Kuldas S, Ismail HN, Hashim S, Bakar ZA. Unconscious learning processes: Mental integration of verbal and pictorial instructional materials. SpringerPlus. 2013;2[1]:105-119.

66. Endsley MR. Expertise and Situation Awareness, In Ericsson KA, Charness N, Feltovich PJ, Hoffman RR. (Eds.). The Cambridge Handbook of Expertise and Expert Performance, 2006, 633-652. NY, US, Cambridge University Press.

67. Schmidt HG, Rikers RM. How expertise develops in medicine: knowledge encapsulation and illness script formation. Medical Education. 2007. 41(12):1133-1139.

68. Ilgen JS, Humbert AJ, Kuhn G, Hansen ML, Norman GR, Eva KW, Charlin B, Sherbino J. Assessing diagnostic reasoning: a consensus statement summarizing theory, practice, and future needs. Acad Emerg Med. 2012;19(12):1454-61.

69. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. Acad Med. 1990;65(10):611-621.

70. Kiesewetter J, Ebersbach R, Görlitz A, Holzer M, Fischer MR, Schmidmaier R. Cognitive problem solving patterns of medical students correlate with success in diagnostic case solutions. PLoS One. 2013;8(8):e71486

71. Scott RB, Dienes Z. The conscious, the unconscious, and familiarity. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2008. 34(5):1264-1288.

72. Nendaz M, Raetzo MA, Junod AF, Vu NV. Teaching Diagnostic Skills: Clinical Vignettes or Chief Complaints? Adv Health Sci Educ Theory Pract. 2000;5(1):3-10.

73. Schuwirth L. Is assessment of clinical reasoning still the Holy Grail? Medical Education. 2009. 43(4):298-300.

74. Singh H, Giardina TD, Petersen LA, Smith MW, Paul LW, Dismukes K, Bhagwath G, Thomas EJ. Exploring situational awareness in diagnostic errors in primary care. BMJ Qual Saf. 2012;21(1):30-38.

75. Huwendiek S, Mennin S, Dern P, Ben-David MF, Van Der Vleuten C, Tönshoff B, Nikendei C. Expertise, needs and challenges of medical educators: Results of an international web survey. Medical Teacher. 2010. 32(11):912-918.

76. Kuper A, D'Eon M. Rethinking the basis of medical knowledge. Medical Education. 2011. 45(1):36-43.

77. Wallace AG. Educating tomorrow's doctors: the thing that really matters is that we care. Academic Medicine. 1997. 72(4):253-258.
78. Ziv A, Ben-David S, Ziv M. Simulation based medical education: an opportunity to learn from errors. Medical Teacher. 2005. 27(3):193-199.
79. Cass GK, Crofts JF, Draycott TJ. The use of simulation to teach clinical skills in obstetrics. In Seminars in Perinatology. 2011. 35(2):68-73 WB Saunders
80. Reid WA, Evans P, Duvall E. Medical students' approaches to learning over a full degree programme. Medical Education online. 2012, 17(1):17205.
81. Davis P, Kvern B, Donen N, Andrews E, Nixon O. Evaluation of a problem-based learning workshop using pre- and post-test objective structured clinical examinations and standardized patients. J Contin Educ Health Prof. 2000;20(3):164-70.
82. Daniel M, Makary MA. Medical error - the third leading cause of death in the US. BMJ. 2016. 353:i2139:476636183.
83. McDonald KM, Matesic B, Contopoulos-Ioannidis DG, Lonhart J, Schmidt E, Pineda N, Ioannidis JP. Patient safety strategies targeted at diagnostic errors: a systematic review. Ann Intern Med. 2013. 158(5Part2):381-389.
84. Newman-Toker DE, Pronovost PJ. Diagnostic errors—the next frontier for patient safety. JAMA. 2009. 301(10):1060-1062.
85. Edozien LC. Situational awareness and its application in the delivery suite. Obstet Gynecol. 2015. 125(1):65-69.
86. Fioratou E, Flin R, Glavin R, Patey R. Beyond monitoring: distributed situation awareness in anaesthesia. British Journal of Anaesthesia. 2010. 105(1):83-90.
87. Chamberland, M., Mamede, S., St-Onge, C., Setrakian, J., Bergeron, L., Schmidt, H. "Self-explanation in learning clinical reasoning: the added value of examples and prompts." Medical Education.2015. 49(2):193-202.
88. Zwaan L, Schiff GD, Singh H. Advancing the research agenda for diagnostic error reduction. BMJ Qual Saf. 2013. 22(Suppl 2):ii52-ii57.
89. Wood TJ. Is it time to move beyond errors in clinical reasoning and discuss accuracy? Adv Health Sci Educ Theory Pract. 2014. 19(3):403-407.
90. Harden RM, Stevenson M, Downie WW, Wilsonet GM. Assessment of clinical competence using objective structured examination. Br Med J. 1975. 22(1):447-451.
91. Schijven MP, Reznick RK, ten Cate OT, Grantcharov TP, Regehr G, Satterthwaite L, Thijssen AS, MacRae HM. Transatlantic comparison of the competence of surgeons at the start of their professional career. Br J Surg. 2010. 97(3):443-449.

92. Ross M, Carroll G, Knight J, Chamberlain M, Fothergill-Bourbonnais F, Linton J. Using the OSCE to measure clinical skills performance in nursing. J Adv Nurs. 1988. 13(1):45-56.
93. Chumley HS. What does an OSCE checklist measure. Fam Med. 2008. 40(8):589-591.
94. Watson R, Stimpson A, Topping A, Porock D. Clinical competence assessment in nursing: a systematic review of the literature. J Adv Nurs. 2002;39(5):421-431.
95. Zartman RR, McWhorter AG, Seale NS, Boone WJ. Using OSCE-based evaluation: curricular impact over time. Journal of Dental Education. 2002. 66(12):1323-1330.
96. Martin IG, Stark P, Jolly B. Benefiting from clinical experience: the influence of learning style and clinical experience on performance in an undergraduate objective structured clinical examination. Medical Education. 2000. 34(7):530-534.
97. Gruppen L, Frank JR, Lockyer J, Ross S, Bould MD, Harris P, Bhanji F, Hodges BD, Snell L, Ten Cate O. Toward a research agenda for competency-based medical education. Medical Teacher. 2017. 39(6):623-630.
98. Daley BJ, Torre DM. Concept maps in medical education: an analytical literature review. Medical Education. 2010. 44(5):440-448.
99. Ginsburg S, McIlroy J, Oulanova O, Eva K, Regehr GS. Toward authentic clinical evaluation: pitfalls in the pursuit of competency. Acad Med. 2010. 85(5):780-786.
100. Van Der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. Medical Education. 2005. 39(3):309-317.
101. Durak HI, Caliskan SA, Bor S, Van der Vleuten C. Use of case-based exams as an instructional teaching tool to teach clinical reasoning. Medical Teacher. 2007. 29[6]:e170-174.
102. Gormley G. Summative OSCEs in undergraduate medical education. The Ulster Medical Journal. 2011. 80(3):127-132.
103. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet, 2001. 357(9260):945-949.
104. Fischer MA, Kennedy KM, Durning S, Schijven MP, Ker J, O`Connor P, Doherty E, Kropmans TBJ. Situational Awareness within Objective Structured Clinical Examination (OSCE) Stations in Undergraduate Medical Training - A Literature Search. BMJ Medical Education. 2017. 17:262 DOI 10.1186/s12909-017-1105-y
105. Endsley MR, Garland D. (Eds.). Situational Awareness Analysis and Measurement. 2000. Mahawah, NJ, US: Laurence Erlbaum Assoiciates Publishers.

106. Varkey P, Natt N. The Objective Structured Clinical Examination as an educational tool in patient safety. Joint Commission Journal on Quality and Patient Safety / Joint Commission Resources, 2007. 33(1):48-53.
107. Durning SJ, Artino A, Boulet J, La Rochelle J, Van der Vleuten C, Arze B, Schuwirth L. The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. Medical Teacher. 2012. 3([1):30-37.
108. LaRochelle JS, Dong T, Durning SJ. Preclerkship assessment of clinical skills and clinical reasoning: the longitudinal impact on student performance. Mil Med. 2015. 180(4 Suppl):43-46.
109. Park WB, Kang SH, Lee YS, Myung SJ. Does objective structured clinical examinations score reflect the clinical reasoning ability of medical students? Am J Med Sci. 2015. 350(1):64-67.
110. Sim JH, Abdul Aziz YF, Mansor A, Vijayananthan A, Foong CC, Vadivelu J. Students' performance in the different clinical skills assessed in OSCE: what does it reveal? Med Educ Online. 2015. 20:26185.
111. Volkan K, Simon SR, Baker H, Todres ID. Psychometric Structure of a Comprehensive Objective Structured Clinical Examination: A Factor Analytic Approach. Adv Health Sci Educ Theory Pract. 2004. 9(2):83-92.
112. Stansfield RB, Diponio L, Craig C, Zeller J, Chadd E, Miller J, Monrad S. Assessing musculoskeletal examination skills and diagnostic reasoning of 4th year medical students using a novel objective structured clinical exam. BMC Med Educ. 2016. 16(1):268-274.
113. Durak HI, Caliskan SA, Bor S, Van der Vleuten C. Use of case-based exams as an instructional teaching tool to teach clinical reasoning. Medical Teacher. 2007. 29[6]:e170-174.
114. Lafleur A, Côté L, Leppink J. Influences of OSCE design on students' diagnostic reasoning. Medical Education. 2015. 49(2):203-214.
115. Myung SJ, Kang SH, Phyo SR, Shin JS, Park WB. Effect of enhanced analytic reasoning on diagnostic accuracy: a randomized controlled study. Medical Teacher. 2013. 35(3):248-250.
116. Furmedge DS, Smith LJ, Sturrock A. Developing doctors: what are the attitudes and perceptions of year 1 and 2 medical students towards a new integrated formative objective structured clinical examination? BMC Medical Education. 2016. 16(1):32 DOI 10.1186/s12909-016-0542-3.

117. Flin R. Measuring safety culture in healthcare: A case for accurate diagnosis. Safety Science. 2007. 45(6):653-667.

118. Borleffs JC, Custers EJ, van Gijn J, ten Cate OT. "Clinical reasoning theater": a new approach to clinical reasoning education. Acad Med. 2003. 78[3]:322-325.

119. Alexander EK. Perspective: moving students beyond an organ-based approach when teaching medical interviewing and physical examination skills. Acad Med. 2008. 83(10):906-909.

120. Gruppen LD, Wolf FM, Billi JE. Information gathering and integration as sources of error in diagnostic decision making. Medical Decision Making. 1991. 11(4):233-239.

121. Salinitri FD, O'Connell MB, Garwood CL, Lehr VT, Abdallah K. An objective structured clinical examination to assess problem-based learning. Am J Pharm Educ. 2012. 76(3):Article 44.

122. Haider I, Kahn A, Imam SM, Ajmal F, Khan M, Ayub M. Perceptions of final professional MBBS students and their examiners about objective structured clinical examination [OSCE]: a combined examiner and examinee survey. J Med Sci. 2016. 24(4):206-221.

123. Baker EA, Ledford CH, Fogg L, Way DP, Park YS. The IDEA Assessment Tool: assessing the reporting, diagnostic reasoning, and decision-making skills demonstrated in medical students' hospital admission notes. Teach Learn Med. 2015. 27(2):163-173.

124. Daud-Gallotti RM, Morinaga CV, Arlindo-Rodrigues M, Velasco IT, Martins MA, Tiberio IC. A new method for the assessment of patient safety competencies during a medical school clerkship using an objective structured clinical examination. Clinics. 2011. 66(7):1209-1215.

125. Ashby SE, Snodgrass SH, Rivett DA, Russell T. Factors shaping e-feedback utilization following electronic Objective Structured Clinical Examinations. Nurs Health Sci. 2016. 18(3):362-369.

126. Siddiqui FG. Final year MBBS students' perception for observed structured clinical examination. J Coll Physicians Surg Pak. 2013. 23(1):20-24.

127. Khairy GA. Feasibility and acceptability of objective structured clinical examination (OSCE) for a large number of candidates: experience at a university hospital. Journal of Family and Community Medicine. 2004. 11(2):75-78.

128. Hammad M, Oweis Y, Taha S, Hattar S, Madarati A, Kadim F. Students’ opinions and attitudes after performing a dental OSCE for the first time: a Jordanian experience. J Dent Educ. 2013. 77(1):99-104.

129. Pugh D, Touchie C, Humphrey-Murto S, Wood TJ. The OSCE progress test–Measuring clinical skill development over residency training. Medical Teacher. 2016. 38(2):168-173.
130. Gregory A, Hogg G, Ker J. Innovative teaching in situational awareness. The Clinical Teacher. 2015. 12(5):331-335
131. Stubbings l. C.W.M.A., Nurses' use of situation awareness in decision-making: an integrative review. Journal of Advanced Nursing. 2012. 68(7):1443–1453.
132. Glavin R, Maran N. Integrating human factors into the medical curriculum. Medical Education. 2003. 37(s1):59-64.
133. Wilkinson TJ, Fontaine S, Egan T. Was a breach of examination security unfair in an objective structured clinical examination? A critical incident. Medical Teacher. 2003. 25(1):42-46.
134. Delzell, JE., Chumley, H., Webb, R., Chakrabarti, S., Relan, A. "Information-gathering patterns associated with higher rates of diagnostic error." Advances in Health Sciences Education. 2009. 14(5):697-711.
135. OSCE Skills. OSCE Station Categories; c2016. Available from: http://www.osceskills.com. (Cited on 2016 Sep 02).
136. Geeky Medics. Clinical Skills (OSCE); c2016. Available from: http:// www.geekymedics.com/category/osce/. (Cited on 2016 Sep 02).
137. Geeky Medics – Awesome Revision Site with Video Guides (Great for OSCEs). Reddit. Available from: https://www.reddit.com/r/medicalschool/comments/1m1c64/ geeky_medics_awesome_revision_site_with_video/. (Retrieved on 15 of June 2017).
138. Bazeley P, Jackson K. Qualitative data analysis with NVivo. 2013: Sage Publications Limited. London, England
139. Kropmans TJ, Cosic K. Assessment of 'borderline' clinical competence using generalisability and decision studies in dental skills training. Eur J Dent Educ. 2004. 8(3):127-132.
140. Trejo-Mejía JA, Sánchez-Mendiola M, Méndez-Ramírez I, Martínez-González A. Reliability analysis of the objective structured clinical examination using generalizability theory. Medical Education Online. 2016. 21:10.3402/meo.v21.31650.
141. Shavelson RJ, Webb NM, Rowley GL. Generalizability theory. American Psychologist. 1989. 44(6):922-932.
142. Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. Medical Teacher. 2012. 34(11):960-992.

143. Cardinet J, Johnson S, Pini G. Applying generalizability theory using EduG. 2011. New York. Routledge.
144. Endsley MR. Toward a theory of situation awareness in dynamic systems. Human Factors. 1995. 37(1):32‑64.
145. Schulz CM, Krautheim V, Hackemann A, Kreuzer M, Kochs EF, Wagner KJ. Situation awareness errors in anesthesia and critical care in 200 cases of a critical incident reporting system. BMC Anesthesiol. 2015. 16(1):4 DOI 10.1186/s12871-016-0172-7
146. Parush A, Kramer C, Foster‑Hunt T, Momtahan K, Hunter A, Sohmer B. Communication and team situation awareness in the OR: Implications for augmentative information display. J Biomed Inf. 2011. 44:477‑85.
147. Lowe DJ, Ireland AJ, Ross A, Ker J. Exploring situational awareness in emergency medicine: Developing a shared mental model to enhance training and assessment. Postgrad Med J. 2016. pii: postgradmedj‑2015‑133772.
148. Graafland M, Bemelman WA, Schijven MP. Appraisal of face and content validity of a serious game improving situational awareness in surgical training. Journal of Laparoendoscopic and Advanced Surgical Techniques. 2015. 25(1):43-49.
149. Frere M, Tepper J, Fischer M, Kennedy K, Kropmans T. Measuring situation awareness in medical education objective structured clinical examination guides. Educ Health. 2017. 30:193-197
150. Carline JD, Paauw DS, Thiede KW, Ramsey PG. Factors affecting the reliability of ratings of students' clinical skills in a medicine clerkship. J Gen Intern Med. 1992. 7(5):506-510.
151. Downing SM. Reliability: on the reproducibility of assessment data. Med Educ. 2004. 38(9):1006-1012.
152. Singh H, Sittig DF. Advancing the science of measurement of diagnostic errors in healthcare: the Safer Dx framework. BMJ Qual Saf. 2015. 24(2):103-110.
153. Lomis KD, Russell RG, Davidson MA, Fleming AE, Pettepher CC, Cutrer WB, Fleming GM, Miller BM. Competency milestones for medical students: Design, implementation, and analysis at one medical school. Medical Teacher. 2017. 39(5):494-504.
154. Wilkinson TJ, Frampton CM, Thompson-Fawcett M, Egan T. Objectivity in Objective Structured Clinical Examinations: Checklists Are No Substitute for Examiner Commitment. Academic Medicine. 2003. 78(2):219-223.

155. Endsley MR. Situation awareness misconceptions and misunderstandings. J Cogn Eng Decis Making. 2015.9:4‑32

156. Yule S, Flin R, Paterson-Brown S, Maran N. Non-technical skills for surgeons in the operating room: A review of the literature. Surgery. 2006. 139(2):140-149.

157. Sandars J, Bax N, Mayer D, Wass V, Vickers R. Educating undergraduate medical students about patient safety: priority areas for curriculum development. Medical Teacher. 2007. 29(1):60-61.

158. Akaike M, Fukutomi M, Nagamune M, Fujimoto A, Tsuji A, Ishida K, Iwata T. Simulation-based medical education in clinical skills laboratory. J Med Invest. 2012. 59(1-2):28-35.

159. Bolstad CA, and Cuevas HM. Integrating Situation Awareness Assessment Into Test and Evaluation. ITEA Journal, 2010. 31 (Retrieved from http://commons.erau.edu/publication/109).

7.7. Dissemination and Other Achievements During the Project

7.7.1. Abstract submissions

The 9th INMED Annual Scientific Meeting held at the Belfast Hilton hotel on July 5th and 6th 2016.

Title: Assessment of Situational Awareness in Undergraduate Medical Education by Objective Structured Clinical Evaluation: A Literature Review

The 10th INMED Annual Scientific Meeting held at the Royal College of Surgeons in Ireland (RCSI) in Dublin at 27th February to 1st March 2017

Title: Assessing Elements of Situation Awareness within OSCE Stations
(withdrawn due to personal circumstances)

The 11th INMED Annual Scientific Meeting held at the Brookfield Health Science Complex in University College Cork from February 7th to 9th 2018.

Title: Evaluation of Consecutive Guided Training to Improve Interrater Agreement in Identifying Elements of Situation Awareness in Objective Structured Clinical Examination Assessments

7.7.2. Oral Presentations

The 9th INMED Annual Scientific Meeting held at the Belfast Hilton hotel on July 5th and 6th 2016.

Title: Assessment of Situational Awareness in Undergraduate Medical Education by Objective Structured Clinical Evaluation: A Literature Review

Lecture to Medical Undergraduate Students at the National University Ireland Galway on Friday the 10th of March 2017.

Title: Situation Awareness in Medical Education

The 11th INMED Annual Scientific Meeting held at University College in Cork on February the 8th 2018

Title: Evaluation of Consecutive Guided Training to Improve Interrater agreement in Identifying Elements of Situation awareness within Objective Structured examination Assessments

Invited presentation at the IDFEA symposium at the Royal College of Surgeons in Dublin on 2nd of May 2018

Title: Development and Evaluation of an Assessment Tool for Identification of Elements of Situation Awareness in OSCEs in Undergraduate Medical Education

7.7.3. Publications

Frere M, Tepper J, Fischer M, Kennedy K, Kropmans T. Measuring situation awareness in medical education objective structured clinical examination guides. Educ Health 2017;30:193-7

Fischer MA, Kennedy KM, Durning S, Schijven MP, Ker J, O`Connor P, Doherty E, Kropmans TBJ. Situational Awareness within Objective Structured Clinical Examination (OSCE) Stations in Undergraduate Medical Training - A Literature Search. BMJ Medical Education, 2017. 17:262 DOI 10.1186/s12909-017-1105-y

Appendix 1

Prisma checklist for literature search

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Situational Awareness within Objective Structured Clinical Examination (OSCE) Stations in Undergraduate Medical Training – A Literature Review	1
ABSTRACT			
Structured summary	2	**Background**: Despite the implementation of innovative methods into medical education curricula, deficient accuracy in diagnostic and therapeutic decision making is continued to be reported. Situational Awareness (SA) has been identified as crucial prerequisite for excellent diagnostic reasoning and clinical judgement. Recent studies suggest that students have little insight into cognitive information processing and utilisation of elements of SA in clinical scenarios and thus, they may not be able to identify essential steps for clinical reasoning. The Objective Structured Clinical Examination (OSCE) may be a suitable instrument to assess students' utilisation of elements of SA in their clinical reasoning. These factors raise the question as to whether or not aviation-like SA training should be purposefully reflected in medical education and assessment. **Objectives**: The purpose of this literature review is to identify elements of SA from undergraduate medical education curricula that could be assessed within OSCE stations based on Endsley´s model of SA. **Methods:** A literature search was carried out for peer-reviewed papers published between January 1975 and February 2017. Selected databases included PUBMED, EMBASE, SCOPUS and PSYCHINFO. Search terms included "objective structured clinical examination", "OSCE" or "objective structured clinical assessment" combined, using the Boolean operator AND, with "non-technical skills", "sense-making", "clinical reasoning", "perception", "comprehension", "projection", "situation awareness", "situational awareness" and "situation assessment. Extraction of publications relating to OSCE assessment in undergraduate medical training identifying elements of clinical practice can be categorised and assigned to the individual levels of SA based on Endsley's model. **Synthesis of Results:** Identified studies described elements of clinical practice within OSCE assessments in undergraduate medical training which were classified to the appropriate level of SA based on Endsley's model and subsequently assigned to the associated subdomain in the self-developed SA assessment tool. Furthermore, studies were categorised either as an assessment tool or an educational tool for utilisation of SA. **Key Findings:** The review identified that OSCEs have the potential to evaluate student's utilisation of elements of SA as part of their clinical reasoning. **Conclusion and Implications:** Assessment of elements of SA as adapted from the model by Endsley might have the potential to be translated into certain aspects of clinical reasoning evaluation using OSCEs.	1 2

		Incorporating the quantitation of utilisation of SA within OSCEs during undergraduate medical training could develop and strengthen teaching on information gathering and efficient processing. If they address the levels of SA, these OSCEs can provide supportive feedback and strengthen educational measures associated with higher diagnostic accuracy and reasoning abilities. **Limitations** include the use of a self-developed classification tool as a result of the absence of an approved existing assessment tool for SA within clinical practice.	2
INTRODUCTION			
Rationale	3	Diagnostic errors or inaccurate selection of treatment modalities are detrimental to patient safety. Despite the implementation of innovative methods, such as simulation-based learning (SBL) and problem-based learning (PBL), into medical education curricula [1; 2; 3], diagnostic and treatment errors are continually reported in the healthcare environment [4;5]. In contrast, aviation is suggested as having a tremendous safety profile based on the reduction of adverse events and fatal accidents over the past decades [6]. An enhanced acknowledgement of SA is attributed to the reduction of critical events in aviation and has subsequently been implemented into other high-risk environments. SA was described by Endsley in respect to aviation as "a person´s mental model of the world around them" [7]. With regard to the clinical practice, SA is essential for recognising the clinical symptoms and signs of a patients´ illness, thereby enabling accurate clinical decision making with respect to diagnosis and best treatment options [8;9]. Deficient or inadequate SA amongst healthcare professionals has been identified by the WHO as a primary parameter in diagnostic errors attributed to human factors [10]. These factors raise the question as to whether or not aviation-like SA training should be purposefully reflected in medical education and assessment. OSCEs may be a suitable instrument to assess students' use of SA in their clinical reasoning. 1. Ziv, A., S. Ben-David, and M. Ziv, Simulation based medical education: an opportunity to learn from errors. Med Teach, 2005. 27(3): p. 193-9. 2. Reid, W.A., P. Evans, and E. Duvall, Medical students' approaches to learning over a full degree programme. 2012, 2012. 17. 3. Davis, P., et al., Evaluation of a problem-based learning workshop using pre- and post-test objective structured clinical examinations and standardized patients. J Contin Educ Health Prof, 2000. 20(3): p. 164-70. 4. Makary, M.A. and M. Daniel, Medical error-the third leading cause of death in the US. BMJ, 2016. 353: p. i2139. 5. McDonald, K.M., et al., Patient safety strategies targeted at diagnostic errors: a systematic review. Annals of internal medicine, 2013. 158(5_Part_2): p. 381-389. 6. EASA Report 2017; https://www.easa.europa.eu/document-library/general-publications/annual-safety-review-2017 7. Endsley, M.R., Situation awareness in aviation systems. Handbook of aviation human factors, 1999: p. 257-276 8. Fore, A.M. and G.L. Sculli, A concept analysis of situational awareness in nursing. Journal of Advanced Nursing, 2013. 69(12): p. 2613-2621. 9. Graafland, M., et al., Training situational awareness to reduce surgical errors in the operating room. British Journal of Surgery, 2015. 102(1): p. 16-23. 10. WHO., Human Factors in Patient safety Review of Topic and Pools. Report for Methods and Measures Working Group of WHO Patient Safety, 2009.	2 - 6

Objectives	4	The purpose of this literature review is to identify whether levels of SA can be assessed during undergraduate medical education curricula utilising OSCEs based on Endsley's model.	6
METHODS			
Protocol and registration	5	Literature search pertaining to papers that describe the assessment of CR using OSCEs among undergraduate medical students. Search terms include "Objective Structured Clinical Examination", "OSCE" or Objective Structured Clinical Assessment" and "non-technical skills", "sense-making", "clinical reasoning", "perception", "comprehension", "projection", "situation awareness", "situational awareness" and "situation assessment". Boolean operators (AND, OR) were used as conjunctions to narrow the search strategy, resulting in the limitation of papers relevant to the research interest. Data were extract and managed for each of the studies included in the final assessment using a structured data recording form. Data processing included information such as type and description of measurement, type of outcome of measurement, assessment of category of SA, type of participants (year of study) and intervention and associated outcome.	6 - 7
Eligibility criteria	6	Publications relating to OSCE assessments within undergraduate medical training published between January 1975 (first paper describing an OSCE) and February 2017, in peer reviewed international journals published in English language were included. Areas of particular interest were elements of SA within OSCEs and the assessment of utilisation of SA within these examinations. Eligibility assessment was performed based on the title, abstract and full text.	6
Information sources	7	Selected databases are PUBMED, EMBASE, SCOPUS and PSYCHINFO.	6
Search	8	The search strategy included filters for the setting as well as terms used in high-risk environments when assessing SA. The MEDLINE search strategy was used for the other databases. For any potential difficulties, an experienced librarian was available to provide assistance in the development of the search terms and searching the electronic databases. The search strategy was limited to studies published between January 1975 and February 2017. Citations referenced in studies included in the final assessment process were retrieved and critically assessed for eligibility for inclusion. Search terms included "Objective Structured Clinical Examination", "OSCE" or "Objective Structured Clinical Assessment" and "non-technical skills", "sense-making", "clinical reasoning", "perception", "comprehension", "projection", "situation awareness", "situational awareness" and "situation assessment". Boolean operators (AND, OR) were used as conjunctions to narrow the search strategy, resulting in the limitation of papers relevant to the research interest.	6
Study selection	9	Any study (randomised or non-randomised design) relating to OSCE and undergraduate medical training and 'situational awareness' or information processing as part of clinical reasoning were included. Areas of particular interest were elements of SA within OSCEs and the assessment within these examinations. Publications relating to nursing, paramedic, pharmacy and veterinary education were excluded from the search. A manual review of the references listed in the identified publications was carried out and any publications of potential interested were sourced and reviewed for eligibility for inclusion.	6

Data collection process	10	The principal investigator (PI) was responsible for selecting studies through the screening, eligibility and inclusion stages. The PI carried out data collection, initially with a pilot extraction form which will be tested and modified for use in final data collection. Upon removal of duplicates the headlines of all publications identified through the search strategy were appraised for eligibility, and either discarded or assessed at abstract for further inclusion. Subsequent exclusion of studies was guided by full-text reviews. Data were extracted and managed for each of the studies included in the final assessment using a structured data recording form. Data processing included information such as type of measurement instrument, description of measurement, type of outcome of measurement, assessment of category of SA, type of participants (year of study, age, gender), intervention and associated outcome.	6
Data items	11	Data were extracted from each included paper based on the identification of elements of medical/ clinical practice which can be assigned to the individual levels of Situational Awareness (SA) and associated subdomains previously stated in the assessment tool for SA. Four key elements of clinical practice were categorised for each level of SA adapted from Endsley´s model for high-risk environments. Level 1 SA: Perception of situational elements included the overall general impression, history taking, physical examination and the retrieval of diagnostic test results. Level 2 SA: Comprehension of elements in situation incorporated pattern recognition, detection of abnormalities within the gathered information, formulating a working diagnosis and the consideration of differential diagnosis. Level 2 SA: Projection of their meaning for future situation included consideration of treatment options, identification of the need for further investigations, consideration of optional outcomes and the search for expedient additional information for a better understanding of the situation.	
Risk of bias in individual studies	12	Researchers can be biased when independently assigning elements of clinical practice assessed in OSCEs to the individual Levels of SA (SA 1, SA 2 and SA3). A clear, structured approach was used to limit the potential of bias. Disagreements were resolved by discussion with the potential to include a field expert as trouble-shooter. No bias assessment tools were used when carrying out the literature review.	
Summary measures	13	The summary of the kind of measurement and of the measurement outcomes provided information categorised into the identification of elements of clinical practice assessed in OSCEs which can be assigned to the individual levels of SA adapted from Endsley´s model as primary outcome measures. Secondary outcomes indicated if OSCEs can be utilised to assess medical students' utilisation of elements of SA as part of their clinical reasoning or if OSCES offer educational properties to teach students how to use SA in their clinical practice.	
Synthesis of results	14	Studies were critically appraised with regard to study design and intervention characteristics. Identified elements of clinical practice within the OSCE assessment were classified for each individual level of SA based on Endsley´s model and subsequently categorised into the subdomains associated with the selected level of SA. Furthermore, results were synthesised for the two categories - Elements of SA which can be assessed in OSCEs as part of clinical reasoning	

		- The OSCE as an educational tool for utilisation of elements of SA	
Risk of bias across studies	15	Studies were found to be very ‘individual’ and unrelated to each other, thus there was no risk of ‘cross over bias between the studies listed. No bias assessment tools were used when carrying out the literature review.	
Additional analyses	16	Manual review of the references listed in the selected papers eligible for the study.	
RESULTS			
Study selection	17	The initial search of the literature retrieved 1127 publications. Upon removal of duplicates 769 publications were left for inclusion into the search. Out of this pool, 642 papers were excluded on review of their titles as seen as being irrelevant to undergraduate medical education. Out of the remaining 127 publications, 98 were removed on review of the abstract as seen as being irrelevant to the research. Out of the remaining pool of 29 papers,18 were removed based on the review of full text as seen as being irrelevant to elements of SA in OSCEs. 11 articles were eligible for inclusion as related to the assessment of elements of SA in undergraduate medical students in OSCEs. Flow diagram attached.	7
Study characteristics	18	(see table below)	8

Author / Year of Publication	Year of Study	Number of Students	Level of Education	SA Level 1	SA Level 2	SA Level 3	Feedback	Assessment Tool for SA	Educational Tool for SA	Research Interest
Volkan 2004	1999	169	year three	History taking, physical examination	Differential diagnosis	Consideration of treatment options		X		Factor analysis of OSCE constructs
Durak 2007	2000-2001	382	year six	Overall impression, history taking, diagnostic test results	Differential diagnosis	Consideration of treatment options, identification the need for further investigations	X		X	Case-based stationary examination
Varkey 2007	2003	42	year three	History taking	Identification of root cause of error		X		X	Root-cause analysis of error
Durning 2012	2010	170	year two	History taking, physical examination	Differential diagnosis			X		Feasibility, reliability, and validity of the evaluation of clinical

										reasoning utilising OSCEs
Myung 2013	2011	145	year four	Physical examination	Differential diagnosis				X	OSCE evaluation impact of pre-encounter analytical reasoning training
Lafleur 2015	2013	40	year five	Physical examination	Diagnostic reasoning				X	Influence of OSCE design on diagnostic reasoning
LaRochelle 2015	2009-2011	514	year four	History taking, physical examination	Clinical reasoning			X		Impact of pre-clerkship clinical reasoning training
Park 2015	2011	65	year four	Overall impression, history taking, physical examination	Differential diagnosis			X		Comparison of clinical reasoning scores and diagnostic accuracy
Sim 2015	2013	185	year five	History taking, physical examination	Data interpretation, clinical reasoning			X		Assessment of different clinical skills using OSCE
Stansfield 2016	2012	45	year four	Physical examination	Diagnostic reasoning			X		Evaluation of embedding clinical examination results into diagnostic reasoning
Furmedge 2016	2013/2014	1280	year one/two	Information gathering	Predefined focus on integration of basic and clinical science		X		X	Acceptability and educational impact of OSCEs in early years

Risk of bias within studies	19	Not identified	
Results of individual studies	20	In a study by Durning based on three successive stations, students were asked to take a history from a patient, synthesise the data and provide the most likely diagnosis and a problem list. Durning, S.J., et al., The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. Med Teach, 2012. 34(1): p. 30-7. La Rochelle detected the potential of OSCEs to identify and foster those students who are experiencing difficulties with diagnostic reasoning and so possibly to prevent problems in subsequent clinical performance. LaRochelle, J.S., T. Dong, and S.J. Durning, Pre-clerkship assessment of clinical skills and clinical reasoning: the longitudinal impact on student performance. Mil Med, 2015. 180(4 Suppl): p. 43-6. Park demonstrated the inability of OSCE scores to correlate with clinical reasoning abilities. However, scores achieved in clinical reasoning OSCEs strongly correlated with diagnostic accuracy.	8 - 10

		Park, W.B., et al., Does objective structured clinical examinations score reflect the clinical reasoning ability of medical students? American Journal of the Medical Sciences, 2015. 350(1): p. 64-67. Sim demonstrated, that out of six evaluation criteria, procedural skills were identified as strongest and clinical reasoning abilities as weakest. He suggested that the low mean scores could be the result of students‘ lack of biomedical knowledge, their inability to incorporate the collected information into the clinical presentation of the patient or a combination of both. Sim, J.H., et al., Students' performance in the different clinical skills assessed in OSCE: what does it reveal? Med Educ Online, 2015. 20: p. 26185. Volkan suggested two fundamental structures for OSCEs. Information gathering was represented by history-taking and physical examination, whereas reasoning and dissemination included hypothetico-deductive testing and differential diagnostic thinking. Based on the findings he highlighted the importance of comprehensive OSCEs to assess the ability to apply both processes simultaneously. Volkan, K., et al., Psychometric Structure of a Comprehensive Objective Structured Clinical Examination: A Factor Analytic Approach. Advances in Health Sciences Education, 2004. 9(2): p. 83-92. Stansfield identified a discrepancy between integrating acquired knowledge into the selected physical examination. Fewer deficits in employing adequate physical examination skills were seen in students who were able to embed their findings into the clinical reasoning process. Stansfield, R.B., et al., Assessing musculoskeletal examination skills and diagnostic reasoning of 4th year medical students using a novel objective structured clinical exam. BMC Medical Education, 2016. 16(1): p. 268. Durak described a model in which hybrid forms of OSCE stations were applied. Based on patient scenarios, students were asked to develop a treatment plan upon the collection of relevant data from history-taking, evaluating signs and symptoms, and the identification of underlying pathophysiological changes. After identifying the most likely diagnosis, students were probed to extract relevant information from the clinical notes and diagnostic results. Durak, H.I., et al., Use of case-based exams as an instructional teaching tool to teach clinical reasoning. Med Teach, 2007. 29(6): p. e170-4. Lafleur observed that students apply more diagnostic reasoning when studying for whole task OSCEs rather than those that focused purely on physical examinations. Backward and forward associations, that is, either looking for evidence to support a suspected diagnosis or the aggregation of all identified symptoms and signs to conclude a diagnosis respectively, are both tasks that demand higher cognitive processing activities and, were strengthened when studying collaboratively for comprehensive OSCEs. Lafleur, A., L. Côté, and J. Leppink, Influences of OSCE design on students' diagnostic reasoning. Medical Education, 2015. 49(2): p. 203-214. Myung compared analytical reasoning ability and diagnostic accuracy by analysing of two groups of students, one of which had received prior education on analytical reasoning and one of which had not. While OSCE scores achieved in both cohorts demonstrated no difference for information gathering, higher diagnostic accuracy was seen in that group of students which had received training in analytical reasoning strategies.	

		Myung, S.J., et al., Effect of enhanced analytic reasoning on diagnostic accuracy: a randomized controlled study. Med Teach, 2013. 35(3): p. 248-50. Varkey suggests that the OSCE was an ideal tool for assessing and teaching non-cognitive skills including SA. In her study, students were asked to identify pivotal information in an error-induced patient encounter. Students were tested in their ability to gather relevant history and to identify the colleague who was accountable for the mistake. Varkey, P. and N. Natt, The Objective Structured Clinical Examination as an educational tool in patient safety. Joint Commission journal on quality and patient safety / Joint Commission Resources, 2007. 33(1): p. 48-53. Furmedge interrogated the appreciation of students for a novel, formative OSCE. The clinical scenario was designed to enable testees to exemplify the integration of skills and knowledge into the understanding of a situation rather than the pure retrieval of recited theoretical knowledge. OSCEs were seen as a learning environment to develop thinking strategies when exposed to clinical scenarios mirroring reality. Furmedge, D.S., L.-J. Smith, and A. Sturrock, Developing doctors: what are the attitudes and perceptions of year 1 and 2 medical students towards a new integrated formative objective structured clinical examination? BMC medical education, 2016. 16(1): p. 1.	
Synthesis of results	21	Evaluation of elements of SA Level 1 were identified in 11 publications, mostly seen in elements such as physical examinations, history taking but also in obtaining an overall impression of the patient and the retrieval of diagnostic test results. All 11 studies demonstrated continuative evaluation of elements of SA Level 2, demonstrated by the integration of the gathered parameters in SA Level 1 into further information processing steps. Only two studies assessed the selection process of optional diagnostic and treatment modalities categorised in SA Level 3. Six papers described the OSCE as having the potential to be an assessment tool for clinical reasoning, a method that might correspond with those used for the assessment of SA in high-risk environments or simulation scenarios. Furthermore, five papers suggested the OSCE as a valuable means for educating medical students on information gathering when they are assessing the identification of the clinical presentation and incorporating the findings into their decision tree.	8
Risk of bias across studies	22	Not identifiable	
Additional analysis	23	Not applicable	
DISCUSSION			
Summary of evidence	24	Six studies concluded that OSCE stations allow for the assessment of students' utilisation of elements of SA as part of their clinical reasoning. Five research studies identified the potential for OSCE stations to be teaching tools for SA within medical education.	

Limitations	25	Limitations include use of a self-developed classification tool as a result of the absence of an approved existing assessment tool for SA within clinical practice.	
Conclusions	26	Assessment of elements of SA as described in the model by Endsley could have the potential to be translated into certain aspects of clinical reasoning assessment using OSCEs. We suggest that students could be exposed to the concept of SA at the early stages in their training and in a simulated format, prior to meeting complex challenging clinical situations in their later medical careers. Efforts in conveying underlying aspects of obtaining and maintaining adequate SA during undergraduate education adapted to their knowledge and expertise could be reflected in enhanced abilities to read and understand clinical scenarios in subsequent postgraduate assessments and clinical practice. Assessment of elements of SA as adapted from the model by Endsley might have the potential to be translated into certain aspects of clinical reasoning evaluation using OSCEs. Given that assessment is a fundamental driver of adult learning, incorporating the quantitation of utilisation of SA within OSCEs during undergraduate medical training could develop and strengthen teaching on information gathering and efficient processing. However, further research needs to establish whether different levels of SA can be identified throughout the medical curriculum and its assessment including the use of paper cases and reviewing medical records. If so, are these levels of assessment congruent with the learning outcomes in preclinical and clinical years? In order to teach students how to perceive and incorporate relevant data, it is essential to provide focussed and informative feedback related to each level of SA and the associated steps of clinical reasoning. Upon identification of the potential and ability to assess levels of SA in a curriculum e. g. OSCEs, we suggest that students be exposed to the concept of SA at the early stages in their training, prior to meeting complex challenging clinical situations in their later medical careers. Efforts in conveying underlying elements of SA during undergraduate education could be reflected in enhanced abilities to read and understand clinical scenarios in subsequent clinical practice.	13
FUNDING			
Funding	27	No funding was received	