Vol.6 Special Issue No.1 (2024)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Does simulation in medical education enhance or inhibit the development of self-knowledge?

Kevin Stirling[A] A *Lead for Simulation and Clinical Skills, School of Health Sciences, University of Dundee, Dundee, Scotland, UK*

Angela Roger[B] B *Senior Lecturer in Education (retired), University of Dundee*

Keith Topping[C] C *Education and Society, University of Dundee & Institute for Foreign Languages, RUDN University, Moscow*

Keywords

Assessment;
medical students;
self-knowledge;
ward simulation exercise.

Correspondence

k.j.stirling@dundee.ac.uk [A]

Article Info

Received 7 July 2023
Received in revised form 28 November 2023
Accepted 19 January 2024
Available online 31 January 2024

DOI: https://doi.org/10.37074/jalt.2023.6.S1.11

Abstract

Simulation has been widely adopted in medical education. Traditionally, the design of simulation activities was through a hierarchical approach where experts within a speciality contributed to the development of content and assessment processes. Whilst this has proved to be a somewhat reliable method, the effectiveness from the perspective of medical students has rarely been examined.

The Ward Simulation Exercise was delivered in the final year of the undergraduate medical curriculum at the University of Dundee. It was designed to assess the capabilities of medical students to prioritise competing demands and work collaboratively within a simulated environment. Medical students were observed by two assessors (normally consultants), who determined whether the student had met the required standard to pass this assessment.

This study examined whether the Ward Simulation Exercise enhanced or inhibited the development of medical students' self-knowledge. This study presented a longitudinal analysis over five years which examined the effect that the Ward Simulation Exercise had on the development of students' self-knowledge. Medical professionals arguably need to be more inclusive of students in designing simulation activities and the associated assessment process. This could allow students to develop self-knowledge appropriate to their stage of professional development.

Introduction

Within medical education, there is a requirement for universities to deliver curricula which ensure that medical students can demonstrate competency in all learning outcomes at the point of graduation (General Medical Council (GMC), 2015; GMC, 2018). In the United Kingdom (UK), the academic journey to graduating as a newly qualified doctor normally takes five years. The construct of medical undergraduate curricula in the UK historically adhered to a positivist paradigm, which facilitates objective measurements of performance and the integration of a spiral curriculum, affording students the opportunity to revisit core concepts and theories and to apply this learning within different clinical specialities (Acton, 1951; Bruner, 1960).

The most obvious evidence of continued adherence to a positivist paradigm within medical education is observed in the way competency is assessed. The assessment of competency in medical education is designed to determine whether a junior trainee or medical student has attained the objectives of a course or a programme of study and whether they have met the required standard to be deemed safe to practise within clinical environments (Tejinder, 2022). The recurring elements of this assessment process are noted to be a need for a junior trainee or medical student to be examined by a person (normally of a superior status) to determine their competency in a skill or a task (Epstein, 2007). Owens (2012) identified that in the year 1027, the imperial physician Wang Wei-Yi had two bronze statues made to assess the competency of his trainees in correctly inserting acupuncture needles. Harden and Gleeson (1975) developed the Observed Structured Clinical Examination (OSCE) framework. The OSCE examination, where senior medical professionals assess the competence of students to complete set tasks specific within a number of stations, has been widely adopted within undergraduate curricula globally. Within clinical practice, the Mini-Clinical Evaluation Exercise (Mini-CEX) and the Direct Observation of Procedural Skills (DOPS) have become popular in assessing the competency of junior trainees or medical students (Alves de Lima et al., 2007; Tsui et al., 2013; Watson et al., 2014; Eggleton et al., 2016). The determination of competency has changed very little over time which has led to medical education being grounded in the 19th Century rather than the 21st Century (McGaghie, 2015). A scoping review undertaken by Long et al. (2022) analysed literature, published between 2000 and 2020, related to the assessment of competency in medical education. An initial return of 4870 articles was reduced to 80 pieces of literature after applying the inclusion criteria. The authors identified that the relevance of the assessment process to the student, the assessor's perceived competence and the context of the assessment influenced how students perceive the credibility of an assessment and its associated feedback. These findings related to assessments conducted within simulated environments (such as the OSCE) or clinical environments (using Mini-CEX or DOPS).

At the University of Dundee, the undergraduate medical curriculum had two elements: a theory-based curriculum in years one to three and an experienced-based curriculum in years four and five. The OSCE was used to assess competency in both elements of the programme across all five years. Mini-CEX and DOPS were used in the experience-based curriculum in years four and five to determine competency in clinical environments. Although a progressive programme of simulation and assessment was integrated into all five years of the undergraduate medical curriculum, the Ward Simulation Exercise (WSE) was one of the only instances where students practised solely within a simulated environment and self-assessment formed part of the assessment process.

The WSE was designed to assess the preparedness of medical students to progress to graduation and was delivered in the final year of the undergraduate programme (Stirling et al., 2012; Till et al., 2015). The WSE was designed to recreate instances from clinical practice within an authentic simulated environment and lasted for 20 minutes (Ker et al., 2006; McIlwaine et al., 2007). Six exercises were developed for the WSE; three were used for the initial run and three for the second run. Students were randomly assigned to an exercise to minimise contamination. Each exercise was developed and validated through a process of shadowing doctors within clinical areas, focus group sessions with healthcare teams, and testing of the simulation before deployment. This ensured that the WSE was delivered in a standardised manner which was representative of clinical practice.

During the WSE, medical students were expected to prioritise the care that three simulated patients (who were admitted to the simulated ward) would receive whilst dealing with timed interruptions and working collaboratively with a qualified nurse. Simulated patients used a script to portray a new admission, a communication scenario or a patient who became acutely unwell. Protocols and guidelines related to characters, which simulated patients would portray, were updated on an annual basis to ensure the WSE was concurrent with changes in clinical practice.

The competencies of students were assessed by a minimum of two senior doctors (most normally consultants or general practitioners). Assessors used a standardised assessment tool to make judgements about students' capabilities. Following the WSE, students conducted a self-assessment whilst watching a video recording of their WSE. This process then informed a feedback session. In this way, the WSE adhered to a positivist paradigm of assessment as medical students were observed by a person of superior status who determined their competency to undertake a skill or task to the required standard.

There was no one way to pass the WSE. The benchmark for a successful performance was how safe the student's response was to the timed interruptions and the patient presentations. The assessment process relied on the professional expertise of the assessor to determine a student's readiness for clinical practice. Although this process might introduce a degree of subjectivity into the assessment process, the WSE was shown to have a good degree of reliability (α=0.89) (Till et al., 2015).

Literature review

Cassam (2014) defines two forms of self-knowledge being trivial and substantial. Trivial self-knowledge is an in-the-moment appraisal that informs judgements relating to one's motivation, capabilities, and ability to perform to an acceptable standard. The concepts of self-efficacy, self-awareness and self-perception arguably describe the basis for the development of trivial self-knowledge (Govern & Marsch, 2001; Aguirre-Raya et al., 2016; Williams et al., 2017). The term 'trivial' self-knowledge might debase what might have been multiple points of significant learning for a student. Therefore, in the context of this study, the term 'trivial' will be replaced with the term 'momentary' self-knowledge which is still distinguishable from 'substantial' self-knowledge. In relation to the WSE, momentary self-knowledge was observed during this assessment when students were motivated to perform well (self-efficacy), they were able to regulate their performance in the moment (self-awareness) and they were able to react to and deal with competing demands (self-perception). Substantial self-knowledge can be considered to be the aggregation of the insights derived from exposure to internal and external stimuli (momentary assessments of self and the subsequent decisions made).

There is limited published literature which reports how the concept of self-knowledge has been utilised within medical education. Dias Pereira et al. (2015) published one of the few articles that addressed the concept of self-knowledge within medical education. This research study evaluated student learning at the end of an elective module that taught stress management strategies. The authors described the impact that stress had on medical students' psychological resilience (the authors stated that burnout, depression and anxiety are highly prevalent in medicine). The authors examined the development of resilience by analysing how students perceived their reactions to stress before and after a standardised intervention and which coping strategies they incorporated into their practice to improve their stress management skills. Questionnaire data (n=76) was reported under three themes (stress symptoms perceived before and after the intervention, use of stress coping strategies by students after the course and students' perceptions of the course). The data reported was surprisingly one-dimensional; 67% reported fewer symptoms of stress, 76% adopted new coping strategies and 90% recognised that stress management strategies would benefit their colleagues. No clear definition of self-knowledge was described or applied in this study, which was a significant limiting factor. Gardiner (2016) commented on the centrality of self-knowledge as part of a patient consultation. Gardiner stated that self-knowledge is critical in informing clinical reasoning to minimise bias and the comparative classification of patient symptoms. The author suggested that self-knowledge facilitates patient-centred consultations that mitigate the tendency to lapse into unconscious mental processing where diagnoses are made that may be erroneous due to an incomplete investigative approach. The author argued that self-knowledge facilitates a creative space where clinicians can reflect and explore multiple possibilities in relation to making a differential diagnosis (challenging perceived assumptions, feelings, thoughts, and beliefs).

The student voice in medical education is not commonly heard in the design of simulation activities or the associated assessment process (Elliot et al., 2019). This research study utilised the standpoint theory which attests that within any social encounter, there are submissive and dominant parties; these roles are interchangeable and are either consciously or unconsciously engaged with (Kokushkin, 2014; Schumann, 2016). Standpoint theory is grounded in feminism and legitimises the opinion and agency of the submissive (whether that is by gender, role or circumstance) and contrasts this with the opinion of those deemed to be in a dominant position (Jovic et al., 2006; Bleakley, 2013). There is limited published literature about the utilisation of standpoint theory within medical education (Bynum and Artio, 2018; Sharma, 2019). The Standpoint theory was used to analyse the opinion of the student (being in a submissive role) with that of the assessors (perceived as being in a dominant role) and the effect that the construct of the WSE and the associated assessment process had on the ability of students to access stores of self-knowledge.

Healthcare is a highly unpredictable and stressful working environment. The impact that extrinsic and intrinsic stressors have on the performance of individuals and teams and their subsequent ability to function safely within unpredictable and stressful environments has been the subject of significant debate and research within medical education (Reason, 2000). The increased adoption of simulation within healthcare education has been catalysed by the mandate to improve patient outcomes and minimise incidences of avoidable harm. The continued adherence to a positivist paradigm negates an understanding of the lived experience of the learner (Rees, 2013). Cronbach's Alpha is a standardised method of reporting the internal consistency of assessment instruments in medicine such as DOPS, Mini-CEX, the OSCE and the WSE. Both Mini-CEX and DOPS can be completed in a relatively short period of time (10 - 15 minutes) and are designed to assess a student's ability to undertake a singular task (either an examination or a procedural skill). Within controlled environments (such as a simulation centre or having dedicated time for assessments), both Mini-CEX and DOPS achieve acceptable levels of reliability (α=0.85) (Watson et al., 2014; Eggleton et al., 2016). The OSCE is structured using multiple stations that students rotate around, completing tasks specific to each station (Harden & Gleeson, 1975). As with Mini-CEX and DOPS, each station assesses competency in a specific skill or procedure. Brannick et al. (2011) conducted a meta-analysis to determine the reliability of the OSCE in medical education. The authors analysed data across assessment criteria and stations. The authors' meta-analysis reported α=0.78 across assessment criteria and α=0.66 stations (the mean α for OSCEs with less than 10 stations was α=0.56 and α=0.74 with greater than 10 stations).

Self-assessment is a subject that generates a significant amount of debate within medicine. The landmark publication by Linn et al. (1975) advocated that medical students should be encouraged to develop skills with peer- and self-assessment to facilitate more diverse evaluation methods (beyond annual examinations) that incorporate an examination of behaviours and motivations that might reveal personal insecurities or overconfidence. Almost 40

years later, Eva and Regehr (2011) were critical of how the medical profession had failed to integrate self-assessment effectively into the practices of the profession, thereby limiting the utility of this technique to inform ongoing professional development. Davis et al. (2006) noted that the current application of self-assessment within medical education predominantly formed part of formal revalidation processes or high-stakes examinations. The authors identified that there was minimal training and preparation within the profession to support doctors to engage effectively with this process. The authors concluded that there was an inadequate evidence base to determine the ability of doctors to undertake self-assessment effectively.

Within medical education, Kruger and Dunning (1999) identified that students in the top percentile of their cohort succumbed to a false consensus where they decreased their self-assessment scores to be more in keeping with their peers. The reverse was shown in those in the lower percentile who lacked insight into their poor performance and rated themselves higher. Therefore, highly reflective and introspective students will embrace negative self-concepts regarding their practice, whilst those with low reflective abilities will endorse a more favourable evaluation of their abilities (Cooney et al., 2021).

Within medical education, the accuracy of a student's self-assessment to describe their lived experience within an assessment is sometimes based on how accurately this perception correlates with the judgements made by assessors (Domician and Éva, 2017; Kukulski et al., 2021). Although self-assessment is endorsed as part of the process of revalidation in medicine, the further application and adoption of this technique within medical education is lacking.

Methodology

The WSE utilised a standardised approach to the delivery of this assessment, which was as follows:

- Assessors observed a student undertake the WSE and rated their performance independently.
- At the end of the WSE, assessors would make a consensus judgement of the student's performance.
- The student would review a video recording of their WSE and conduct a self-assessment.

After completing their self-assessment, students would receive feedback from the assessors who described whether they had met the required standard or not. Those students who had not met the required standard on their first attempt were invited to undertake a second WSE. This transactional approach to feedback, where the onus is placed on the student to enact change following an assessment process, is commonplace in medical education. A scoping review conducted by Cordovani et al. (2023) identified that the process of providing student feedback was a relatively new development in medical education (it has increased in popularity since 2010). The authors identified a variety of approaches to delivering feedback which varied in methodological rigour. The authors concluded that the ability of students to interpret the feedback provided and the credibility of the person delivering that feedback were core determinants of whether students could integrate this information into their practice. As the WSE was an uncommon assessment, being one of the few times that a medical student practised alone in a simulated environment, this did pose ontological challenges for students regarding how they might assimilate this feedback, which might vary in the overall level of quality and relevance from the student standpoint, successfully into their practice.

The assessment instrument used by students (to conduct a self-assessment) and assessors (to conduct an independent and consensus judgement) during the WSE had a mix of open questions and closed domains. Open questions were descriptive in nature and evaluated:

- the strengths in their practice,
- areas for improvement, and
- how this will affect their clinical practice.

Closed domains rated performance using a 1–- 5 Likert scale (1) Very poor, 2) Poor, 3) Good, 4) Very good and 5) Outstanding) and were aligned with the domains in Good Medical Practice, which was the national guidance framework by which a student's performance would be assessed when they graduated (GMC, 2019). The assessors' independent and consensus forms mirrored the student domains in 2010, except that the global score was replaced with a pass/fail domain. As part of their evaluation of the reliability of the WSE (using data from 2010), Till et al. (2015) identified that two domains, Communication and Health and Safety, were the easiest to score highly in. In 2012, the Communication domain was expanded to include ratings for i) Communication with Patients and ii) Communication with Relatives. The Health and Safety domain was expanded to include ratings for i) Safe Medical Practice and ii) Preventing Cross Infection. A global score was inserted to allow direct comparison with the student's self-assessment. The process of making a consensus judgement was amended in 2014. The Global Score and consensus pass/fail judgement were removed at the request of the deanery, and students were only provided with open text feedback following the WSE. This change was intended to make the consensus feedback more relatable to the student's ongoing professional development. Therefore, the process by which assessors used closed domains to make judgements pertaining to students' performance differed for every year of data collection (the open questions remained the same). This was not ideal for the process of data analysis and reporting. Most domains remained consistent in 2010, 2012, and 2014, but the impact of any amendments will be acknowledged as part of the data analysis process.

The domains from Good Medical Practice which underpinned the closed domains in the assessment instrument were as follows:

- Prioritisation (of essential tasks and procedures)
- Clinical skills (effective technical skills)
- Acutely unwell patient (recognition and systematic assessment)
- Prescribing and written documentation (completion of written tasks and safe prescribing) ^
- Response to interruptions.
- Communication (good interpersonal skills).
  - o With patients/relatives *
  - o With colleagues *
- Health and safety (demonstrating safe practice)
  - o Safe medical practice *
  - o Preventing cross-infection *
- Professionalism *
- Global score (overall rating of the student's performance) *

* These domains were added to the assessment process in 2012.

^ This domain was separated into two separate domains ('Prescribing technique' and 'Written documentation') in 2012.

In relation to the WSE, only data from assessors has been reported in previous publications. This research study distanced itself from the positivist standpoint which had underpinned previous research activity surrounding the WSE and engaged with a post-positivist standpoint to generate new knowledge regarding how simulation enhances or inhibits the development of student self-knowledge.

Assessment data from three years were analysed using a mixed methods approach. Quantitatively, Cronbach's Alpha was used as a measure of the reliability of the assessment tools used as part of the WSE. Qualitatively, K-means Cluster Analysis and Inductive Thematic Analysis were used to better understand the lived experience of students and determine the reliability of the assessment from the student perspective.

Cronbach's alpha

The closed domains from students' self-assessments and assessors' independent and consensus judgement forms for the first and second run of the WSE in 2010, 2012 and 2014 were analysed using Microsoft Excel. The reliability of each assessment instrument was reported for each year and run. An overall statement regarding the reliability of the instrument over multiple years was also made. Cells that had no data recorded had a zero inserted. The total number of empty cells for each year and run was reported as part of this stage of analysis.

K-means cluster analysis

K-means Cluster Analysis categorised data beyond statements and judgments that were directly linked to whether a student met the required standard. The closed domains from students' self-assessments and assessors' independent assessment forms were used to define the number of clusters to be generated as part of the data analysis. Students' self-assessment data from 2010, 2012 and 2014 were assigned to seven clusters. Assessors' independent assessment forms were assigned to eight clusters in 2010 and 2012 and 12 clusters in 2014. K-means Cluster Analysis assigned assessment domains (in the case of the WSE, all scores submitted by students and assessors) into pre-determined groups (clusters) that explained the structure of the data. Clusters are ranked in order of significance and the volume of objects assigned to each group. This allowed a determination of convergence or non-convergence in relation to what both groups determined to be the most important components of a satisfactory performance during the WSE.

Data were analysed using SPSS version 22. K-means Cluster Analysis created a bridge between the quantitative and qualitative methods which informed the stages of Inductive Thematic Analysis.

Inductive thematic analysis

Inductive Thematic Analysis advocates a constant comparative approach when analysing data (Thomas, 2006; Braun & Clarke, 2006). The constant comparative approach appraises data holistically and builds concepts based on what the data say most frequently and with the greatest clarity. This approach requires an exhaustive knowledge of the data and rejects any urge to superficially code data.

Qualitative data was reviewed from the standpoint of a student's self-assessment of their performance during the WSE as follows:

- Open text statements were coded to generate categories.
- Categories were refined through multiple readings.
- Prominent categories were contrasted with open-text data from assessors.
- Convergence or non-convergence (between students and assessors) was then determined, describing instances where the WSE enhanced or inhibited the development of self-knowledge.

Data were generated from responses to the open questions in students' self-assessment forms and assessors' independent and consensus judgement forms from the first and second runs of the WSE in 2014. This systematic approach allowed for multiple categories, properties and hypotheses to be generated that proposed and validated a formal theory that will be discussed further in the results. To ensure that the process of theory generation was robust, data were

reported using the cluster rankings generated by K-means Cluster Analysis. A process of selective secondary review of qualitative data was undertaken by a colleague from the University of Dundee. This colleague has experience in the delivery of the undergraduate medical curriculum including the annual diet of examinations. This colleague was selected as they were aware of the construct of the WSE (including its assessment process) but they had no active role in the delivery of this assessment.

Ethical considerations

Ethical approval for this research study was granted by the University of Dundee Research and Ethics Committee (ref no: UREC15047). Prior to undertaking the WSE, students gave their informed consent to allow the materials relating to their performance to be used for research purposes. This request for consent occurred prior to the preparatory briefing that was delivered just before a student undertook the WSE. Students completed the consent form independently and on a voluntary basis without coercion.

In all instances, any identifiable data relating to the first or second run of the WSE in 2010, 2012 and 2014 were anonymised prior to analysis.

Analysis

The total volume of data sets from the first and second run of the WSE in 2010, 2012 and 2014 is presented in Table 1.

Table 1: Data sets of students and assessors for the first and second run of the WSE.

	1st run of the WSE				2nd run of the WSE			
Year	Data set 1		Data set 2		Data set 3		Data set 4	
2010	Students	158	Assessors	32	Students	29	Assessors	14
	Data set 5		Data set 6		Data set 7		Data set 8	
2012	Students	153	Assessors	33	Students	228	Assessors	19
	Data set 9		Data set 10		Data set 11		Data set 12	
2014	Students	165	Assessors	41	Students	47	Assessors	17

The total amount of data collected from the first and second run of the WSE in 2010, 2012 and 2014 is presented in Table 2.

Table 2: The total amount of data generated in 2010, 2012 and 2014 from the first and second run of the WSE.

Year	Assessors				Students		Total
	Independent		Consensus		Self-assessment		
	1st run	2nd run	1st run	2nd run	1st run	2nd run	
2010	155	45	155	45	155	45	600
2012	154	45	154	45	154	45	597
2014	163	52	-	-	163	52	430
Total	472	142	309	90	472	142	1627

Data were excluded if it failed to meet the following criteria (listed in order of priority):

- The name of the student or assessor must be identifiable.
- An assessment form must have a minimum of 50% of the domain fields completed.
- Assessors and students must complete the global rating score (excluding 2010 data).
- A pass or fail judgement must be made by assessors when completing the independent and consensus assessment forms (excluding 2014).

The inclusion criteria were applied to the assessors' independent and consensus forms, and if no exclusions were identified, then they were applied to the students' self-assessment. In total, 1552/1627 (95%) data sets from the first and second runs of the WSE in 2010, 2012 and 2014 met the inclusion criteria. The process of data reconciliation resulted in 412/472 data sets (87%) for the first run of the WSE and 127/142 data sets (89%) for the second run of the WSE being included in the subsequent data analysis.

Demographics

Demographic data relating to the gender of the student participants were collected as part of this study (Table 3). Demographic data relating to the professional role (Consultant, General Practitioner (GP), Specialty Trainee (ST)) and the gender of the assessors is presented in Table 4. No further demographic data were collected.

Table 3: Demographic data for students categorised by WSE run and gender.

Year	Run	Male	Female
2010	1st	60	81
	2nd	24	18
2012	1st	51	86
	2nd	16	24
2014	1st	57	77
	2nd	23	22

Table 4: Demographic data for assessors categorised by professional role and gender.

Professional role	Gender	Number of Assessors
Consultant	Male	30
	Female	15
GP	Male	4
	Female	4
ST	Male	7
	Female	9

Quantitative analysis

Quantitative analysis determines the reliability of the assessment instrument used during the Ward Simulation Exercise.

Cronbach's alpha

Till et al. (2015) reported a Cronbach's Alpha of 0.89 for the assessment instrument used during the WSE. This statement of reliability is related to the independent judgements made by assessors during the first run of the WSE in 2010. Cronbach's Alpha was performed to determine the reliability of the assessment instrument used to conduct student self-assessments and the independent and consensus judgements made by assessors as part of the first and

second run of the WSE in 2010, 2012 and 2014. A value of 0.7 and higher was adopted to state whether Cronbach's alpha measures were sufficiently consistent to indicate the measure is reliable. The results for the first and second run of the WSE are presented in Tables 5 and 6.

Table 5: Cronbach's Alpha data for all assessment domains for the first run of the WSE.

Year	Number of students	Independent	Consensus	Self-assessment
2010 [1]	141	**0.89**	**0.80**	**0.83**
2012 [2]	137	**0.95**	**0.90**	**0.78**
2014 [3]	134	**0.91**	-	**0.78**
a ([1,2,3])	-	**0.92**	**0.87**	**0.80**

Emboldened = significant

Table 6: Cronbach's Alpha data for all assessment domains for the first run of the WSE.

Year	Number of students	Independent	Consensus	Self-assessment
2010 [1]	42	**0.80**	0.60	**0.80**
2012 [2]	40	**0.91**	**0.85**	**0.81**
2014 [3]	45	**0.93**	-	**0.76**
a ([1,2,3])	-	**0.92**	**0.87**	**0.78**

Emboldened = significant

Tables 5 and 6 demonstrate that the assessment instrument used by students and assessors showed high levels of internal consistency except for the second run of the WSE in 2010. The α reported in Table 5 for the first run of the WSE in 2010 was similar to what was reported by Till et al. (2015). The number of items with no data recorded totalled 2% for assessors' independent forms (239/12184), 1% for assessors' consensus forms (45/3765) and 1% for student self-assessments (34/4312) which was acceptable.

K-means cluster analysis

The gender of the student or assessor was included in the K-means Cluster Analysis. The gender of the student was analysed along with the assessors' independent data to see if the gender of the person being observed affected the assessment outcome. The gender of both assessors was analysed along with students' self-assessment data to see if an awareness of who was observing their practice affected students' rating of their performance (students were informed who their assessors were as part of the preparatory briefing).

A one-way ANOVA was conducted as part of the analysis. The null hypothesis for a one-way ANOVA is that all means are equal (or exhibit minimal variance). A null hypothesis is stated for each stage of analysis and is reported as part of the results for every year and run. Degrees of freedom (df) are calculated in two ways: df1 assumes that if n equals the number of clusters, the degrees of freedom is n – 1, df2 is calculated by subtracting the total amount of clusters from the total population. Means are derived from the degrees of freedom between clusters (df1) and the total variance within clusters (df2). SPSS version 22 reported df2 as an 'error'. This term was changed to 'variance within clusters' to avoid confusion.

The Qualitative aspects of this study analysed open text data from the first and second runs of the WSE in 2014. Therefore, only the results from the K-means Cluster Analysis for the first and second run of the WSE in 2014 are reported in this section. Tables 7 – 10 report the number of cases allocated to each cluster for students and assessors for the first and second runs of the WSE in 2014. Each table reports the results of the one-way ANOVA for each cluster. A statistically significant difference was stated if the significance value was below 0.05. The total number of assessments reported in Tables 8 and 10 is double the number of students who undertook this assessment. This figure is correct as two independent assessments were conducted for each student in the first and second run of the WSE.

First run of the WSE: 2014

Students.

Table 7: Probabilities of difference based on a student's perception of their performance and the number of cases assigned to each cluster following analysis of students' self-assessment during the first run of the WSE in 2014.

Number of students 134	Variance between clusters		Variance within clusters		F	Anova	No. of cases	%
Cluster	Mean Square	df (1)	Mean Square	df (2)				
Prioritisation	4.4	8	.37	125	12	**.00**	9	7
Clinical Skills	4.2	8	.30	125	14	**.00**	3	2
Acutely Unwell Patient	5.4	8	.35	125	15	**.00**	1	1
Prescribing and Written Documentation	7.6	8	.28	125	27	**.00**	**31**	**23**
Response to Interruptions	6.6	8	.33	125	20	**.00**	**20**	**15**
Communication	3.3	8	.33	125	9.9	**.00**	17	13
Health and Safety	8.9	8	.40	125	22	**.00**	15	11
Assessor 1 Gender	.62	8	.23	125	2.7	**.01**	**26**	**19**
Assessor 2 Gender	.61	8	.23	125	2.7	**.01**	12	9
					Total		134	100

Emboldened = significant/ highest number of cases assigned

The clusters (assessment domains) with the most cases allocated to them were Prescribing and Written Documentation (F(1.1,.18) = 6.2 p= <.00), Acutely Unwell Patient (F(3.0,.18) = 17 p= <.00) and Communication (F(1.1,.21) = 4.9 p = <.00).

Assessors.

Table 8: Probabilities of difference based on an assessor's perception of students' performance and the number of cases assigned to each cluster following K-means Cluster Analysis during the first run of the WSE in 2014.

Number of students 134	Variance between clusters		Variance within clusters		F	Anova	No. of cases	%
Cluster	Mean Square	df (1)	Mean Square	df (2)				
Prioritisation	16	12	.32	255	48	**.00**	8	3
Clinical Skills	9.4	12	.43	255	22	**.00**	4	1
Acutely Unwell Patient	16	12	.35	255	46	**.00**	28	10
Prescribing Technique	31	12	.44	255	72	**.00**	8	3
Written Documentation	17	12	.58	255	30	**.00**	**30**	**11**
Response to Interruptions	6.3	12	.32	255	20	**.00**	4	1
Communication: With Patients	7.0	12	.34	255	21	**.00**	26	10
Communication: With Team	7.7	12	.40	255	19	**.00**	29	11
Health and Safety: Safe Practice	31	12	.45	255	68	**.00**	10	4

Health and Safety: Cross Infection	6.7	12	.81	255	8.3	**.00**	1	0
Professionalism	13	12	.33	255	39	**.00**	**47**	**18**
Pass / Fail	3.1	12	.08	255	37	**.00**	**47**	**18**
Student Gender	.71	12	.22	255	3.2	**.00**	26	10
					Total		268	100

Emboldened = significant/ highest number of cases assigned

The clusters (assessment domains) with the most cases allocated to them were Professionalism (F(13,.33) = 39 p= <.00), Pass/ Fail Judgement (F(3.1,.08) = 37 p= <.00) and written documentation (F(17,.58) = 30 p= <.00).

Second run of the WSE: 2014

Students.

Table 9: Probabilities of difference based on a student's perception of their performance and the number of cases assigned to each cluster following analysis of students' self-assessment during the second run of the WSE in 2014.

Number of students 45	Variance between clusters		Variance within clusters		F	Anova.	No. of cases	%
Cluster	Mean Square	df (1)	Mean Square	df (2)				
Prioritisation	1.5	8	.24	36	6.5	**.00**	3	7
Clinical Skills	1.2	8	.22	36	5.7	**.00**	**11**	**24**
Acutely Unwell Patient	1.7	8	.26	36	6.7	**.00**	1	2
Prescribing and Written Documentation	2.0	8	.31	36	6.3	**.00**	1	2
Response to Interruptions	1.0	8	.23	36	4.4	**.00**	**9**	**20**
Communication	3.1	8	.24	36	13	**.00**	**8**	**18**
Health and Safety	2.9	8	.18	36	16	**.00**	1	2
Assessor 1 Gender	.32	8	.19	36	1.7	.14	7	16
Assessor 2 Gender	.51	8	.10	36	4.9	**.00**	4	9
					Total		45	100

Emboldened = significant/ highest number of cases assigned

The clusters (assessment domains) with the most cases allocated to them were Clinical Skills (F(1.2,.22) = 5.7 p= <.00), Response to Interruptions (F(1.0,.23) = 4.4 p= <.00) and Communication (F(3.1,.24) = 13 p = <.00).

Assessors.

Table 10: Probabilities of difference based on an assessor's perception of students' performance and the number of cases assigned to each cluster following K-means Cluster Analysis during the second run of the WSE in 2014.

Number of students 45	Variance between clusters		Variance within clusters		F	Anova	No. of cases	%
Cluster	Mean Square	df (1)	Mean Square	df (2)				
Prioritisation	5.8	12	.27	77	22	**.00**	3	3%
Clinical Skills	5.6	12	.21	77	27	**.00**	3	3%
Acutely Unwell Patient	5.5	12	.23	77	24	**.00**	**18**	**20%**
Prescribing Technique	8.6	12	.31	77	28	**.00**	5	6%
Written Documentation	8.1	12	.28	77	29	**.00**	**9**	**10%**
Response to Interruptions	2.4	12	.35	77	7	**.00**	**14**	**16%**
Communication: With Patients	3.9	12	.33	77	12	**.00**	8	9%
Communication: With Team	2.5	12	.22	77	12	**.00**	2	2%
Health and Safety: Safe Practice	11	12	.52	77	22	**.00**	3	3%
Health and Safety: Cross Infection	5.4	12	.46	77	12	**.00**	5	6%
Professionalism	3.0	12	.26	77	12	**.00**	5	6%
Pass / Fail	1.2	12	.06	77	19	**.00**	8	9%
Student Gender	0.6	12	.20	77	3	**.00**	7	8%
					Total		90	100

Emboldened = significant/ highest number of cases assigned

The clusters (assessment domains) with most cases allocated to them were Acutely Unwell Patient (F(5.5,.23) = 24 p= <.00), Response to Interruptions (F(2.4,.35) = 7 p= <.00) and Written Documentation (F(8.1,.28) = 29 p= <.00).

Summary of the quantitative analysis

In relation to students who undertook the first and second runs of the WSE in 2010, 2012 and 2014, the null hypothesis was that the gender of assessors would have an influence on the manner in which students undertook their self-assessment. In all instances, except the first run of the WSE in 2012, the gender of assessors did not have a strong effect.

In the first run of the WSE, the assessment domains (clusters) which had cases allocated to them most frequently as part of the K-means Cluster Analysis were as follows (the years(s) in which cases were highest for a cluster is shown in parenthesis): Acutely Unwell Patient (2010, 2014), Prescribing and Written Documentation (2010, 2014), Communication (2012, 2014), Clinical Skills (2012), Assessor 1 gender (2012) and Health and Safety (2010). This was similar to the results reported from the ANOVA in each year. In the second run of the WSE Clinical Skills, the clusters (assessment domains) which had cases allocated to them most frequently as part of the K-means Cluster Analysis were as follows: Acutely Unwell Patient (2010, 2012), Prescribing and Written Documentation (2010, 2012), Communication (2012, 2014), Health and Safety (2010)), Clinical Skills (2014) and Response to Interruptions (2014). This was similar to the results reported from the ANOVA in each year.

Overall, the following clusters had cases assigned most frequently - Acutely Unwell Patient (4/6 runs), Communication (4/6 runs) and Prescribing and Written Documentation (4/6 runs).

In relation to assessors, the null hypothesis was that the gender of students would have an influence on the manner in which assessors undertook their independent assessment. The one-way ANOVA reported statistically significant results for all clusters in all years, but student gender, apart from the second run of the WSE in 2010, did not have a strong effect.

In the first run of the WSE in 2010, assessment data demonstrated convergence with students in three domains (Health and Safety, Acutely Unwell Patient and Prescribing and Written Documentation) and non-convergence in one (Clinical Skills). In the second run of the WSE in 2010, assessment data demonstrated convergence with students in all domains (Health and Safety, Acutely Unwell Patient and Prescribing and Written Documentation).

In the first run of the WSE in 2012, assessment data demonstrated convergence with students in one domain (Communication) and non-convergence in two domains (Prioritisation and Prescribing Technique). In the second run of the WSE in 2012, assessment data demonstrated convergence with students in one domain (Written Documentation) and non-convergence in two domains (Pass/ Fail Judgement and Professionalism) which did not

form part of the self-assessment form.

In the first run of the WSE in 2014, assessment data demonstrated convergence with students in one domain (Written Documentation) and non-convergence in two domains (Pass/ Fail Judgement and Professionalism) which did not form part of the self-assessment form. In the second run of the WSE in 2014, assessment data demonstrated convergence with students in one domain (Response to Interruptions) and non-convergence in two domains (Acutely Unwell Patient and Written Documentation).

Overall, the following clusters had cases assigned most frequently - Prescribing and Written Documentation (and its variants) (6/6 runs), Acutely Unwell Patient (2/6 runs), Pass/ Fail Judgement (2/6 runs), Professionalism (2/6 runs) and Response to Interruptions (2/6 runs).

Qualitative analysis

Based on the outputs from the K-means Cluster Analysis, Tables 19 and 20 describe the assessment domain (cluster) rankings which were applied to the process of structuring and analysing the qualitative data:

Table 11: Total number of cases by domain for students and assessors during the first run of the WSE.

First run of the WSE in 2010, 2012 and 2014			
Students		Assessors	
Domain	Total number of cases	Domain	Total number of cases
Gender	98	Communication	84
Communication	65	Prescribing and Written Documentation	64
Prescribing and Written Documentation	65	Prioritisation	47
Clinical Skills	56	-	-

Table 12: Total number of cases by domain for students and assessors during the second run of the WSE.

Second run of the WSE in 2010, 2012 and 2014			
Students		Assessors	
Domain	Total number of cases	Domain	Total number of cases
Prescribing and Written Documentation	27	Prescribing and Written Documentation	22
Clinical Skills	24	Acutely Unwell Patient	17
Gender	21	Response to Interruptions	16
Communication	16	Communication	14
		Gender	14

Clusters were used to categorise and describe personas, activities, interactions and standpoints that related directly to the design and delivery of the WSE. Inductive Thematic Analysis facilitated a more in-depth understanding of why students and assessors were allocated to specific clusters. Analysis of open text data from the first and second runs of the WSE in 2014 described instances of convergence and non-convergence between students and assessors beyond merely contrasting what each professional group deemed acceptable performance components. The framework that sequenced the qualitative analysis was based on the total number of cases assigned to clusters as part of the analysis of student self-assessment data from the first and second run of the WSE in 2014, and were as follows:

1) Gender
2) Communication skills
3) Prescribing and written documentation
4) Clinical skills
5) Prioritisation
6) Acutely unwell patient
7) Response to interruptions

This research study investigated the components (in the design or delivery) of the WSE that enhanced or inhibited the development of self-knowledge. As described in the methodology, this data underwent a process of selective secondary review to ensure that inter-rater reliability was 100%. Codes, concepts and themes from either students or assessors within the text are emboldened.

Students were allocated a unique identifier that was generated by their designation (student (ST)), their gender (male or female (M or F)) and a random allocation of a number to differentiate each data set, for example, STM123. Assessors were allocated a unique identifier that was generated by their designation (Assessor (A)), their status (Consultant (CO), General Practitioner (GP), Speciality Trainee (ST)), their gender (male or female (M or F)), the total number of times they assessed the WSE (01 - 74), for example, ACOM12-01.

Gender

Students

This cluster reported the characteristics of male and female students from the standpoint of the student and thereafter, the assessors. In total, 38 students (13(m), 25(f)) were allocated to this cluster in the first run of the WSE and 11 students (6(m), 5(f)) from the second run.

Male and female students seemed to aspire to inhabit different personas. Overall, the characteristics displayed by male students were more confident and they were focussed on accomplishing tasks, whilst female students were less confident in their abilities and were more collaborative in their approach to teamworking and patient care.

In the first run of the WSE, the persona described by all male students in their self-assessment was one of inspiring confidence, being proactive, knowledgeable and in control. These included undertaking interventions such as **patient consultations** (STM56, STM97), **managing the acutely unwell patient** (STM08, STM126, STM79) and appropriate **patient-centred** and **professional communication** (STM113, STM116, STM43) and **seeking advice from their senior colleague** or to **delegate tasks to the nurse** (STM08, STM43).

The persona described by all female students was predominantly focussed on being person-centred. The characteristics included having a **systematic approach** to care delivery (STF161, STF141, STF110), a **calm demeanour** (STF106, STF98, STF18), **good prioritisation skills** (STF110, STF83, STF77, STF18), a recognition of one's personal limitations (STF158, STF54, STF53), a **reassuring manner towards patients** (STF57, STF09), **collaborative teamworking** (STF137, STF92, STF83, STF53) and **strong patient-centred** and **professional communication** (STF154, STF136, STF129, STF127, STF125, STF83).

In the second run of the WSE, male students were more focussed on being **systematic** and **proactive** (STM145, STM43), ensuring **collaborative teamworking** (STM145, STM105, STM62) and **being patient-centred in practice** (STM105, STM62, STM43). Students described **pausing**, **taking a step back** and **being more organised** as approaches that underpinned the prioritisation of their workload and managing the associated **internal and environmental stressors** (ST14M167, ST14M139, ST14M62).

Similar to the first run of the WSE, female students were focussed on **good communication skills and teamworking** (STF109, STF27, STF25). Students were critical of their ability to **remain calm** and manage internal and external stressors (STF63, STF27), ensure a **systematic approach when assessing patients** (STF109, STF63, STF40, STF25) and **that information was properly communicated and documented** (STF109, STF25).

Assessors

The statements made by assessors (both independently and in consensus) resonated with the personas described by male and female students.

Those male students who passed the WSE attained a standard of practice that was described as consistent, responsive and efficient (ACOF74-01, ACOF46-02, ACOM41-01). These students were observed to be **safe** (ACOM25-01, ASTM37-01), systematic (ACOF74-01), **calm** (ASTM37-01, AGPM46-01) and **organised** (ACOM25-01).

Assessors recognised female students' **communication skills** (ACOM35-01, ACOM25-01, ACOF08-01, ACOM04-01, AGPM07-03, ASTF72-01), **collaborative practice** (ACOM12-03, AGPM04-08), explanation of **management plans** (ASTF60-01, ASTF39-01, ACOM35-01,) and the **delegation of tasks** (ACOM36-01, ACOF08-01, ASTM07-01) as strengths although there was a recognised need to **increase personal confidences** to lead the team more effectively (ACOM25-01, ASTM37-01).

Communication

In total, 17 students (35% (m) and 65% (f)) were allocated to this cluster in the first run of the WSE and 8 students (62% (m) and 38% (f)) from the second run. Due to the volume of data returned, only the first run of the WSE is reported.

In the first run of the WSE, students described limitations in how they **requested investigations** (STM145, STM147, STM82) and **how they conveyed this information to their senior colleagues** (STM145, STM82). When communicating with simulated patients, students recognised that **introducing themselves to the patient** (STM147), **being empathetic** (STM35) and **conducting a structured consultation** (STM35, STM69) as areas of practice which required improvement.

All female students rated their communication skills as a strength. Person-centred communication was described as being **empathetic** (STF74, STF33), **explaining treatment plans** (STF25, STF74) and **conducting an appropriate consultation** (STF90). Areas for improvement were identified as **speaking clearly and concisely to patients** (STF74, STF140), working effectively with the **nurse to deliver interventions** (STF90) and **conducting an effective handover at the end of the WSE** (STF33, STF29, STF146).

Assessors

The management of the acutely unwell patient was a key determinant of whether a male student passed or failed the WSE. Male students who passed the WSE (n=4) **recognised and assessed the acutely unwell patient quickly** (ACOM30-01, ACOF22-01, ASTF72-01, ASTF39-01). They **communicated the results of this assessment** to the nurse or their senior colleague and **worked collaboratively to implement an appropriate treatment plan** (ACOM04-09, ACOF22-01) and to reassess the patient (ACOM30-01).

The language used by assessors to describe female students who passed the WSE (n=7) demonstrated an ability to **remain focused** throughout the WSE. These students **clarified tasks** at the end of the initial handover (AGPM07-03, ASTF39-01), engaged with the **patient's agenda** (AGPM07-03), **conducted a structured consultation** (AGPM07-03, ASTF39-01, ACOM41-01, ACOM04-09) and delivered a **good handover at the end of the WSE** (AGPM46-01).

Prescribing and written documentation

In total, 31 students (52% (m) and 48% (f)) were allocated to this cluster in the first run of the WSE and only one student from the second run. Due to the volume of data returned, only the first run of the WSE is reported. An average score of 2 (poor) was calculated from all self-assessment scores submitted by male and female students.

Students

Of the 16 male students, only one student (STM32) recognised their prescribing technique as a strength. No male students commented positively regarding their documentation skills. All students described instances during the WSE where they felt disorganised and **struggled to manage their workload** (STM10, STM32, STM121, STM135). The underlying cause was described as **not being familiar with the environment or the paperwork** (STM42, STM108), **poor documentation of interventions** (STM68, STM42, STM32) and **an absence of**

a safe prescribing technique (STM132, STM111). The effect of stress on performance was described frequently as a key limiting factor (STM132, STM135).

Like their male counterparts, the 15 female students did not identify any positive aspects of their practice with this domain. The open-text comments of female students described a need to be more **systematic and structured in their prescribing technique** and the **documentation of patient information** (STF01, STF89, STF114). This resulted in data from **patient consultations not being written down** and subsequently forgotten (STF13, STF133) and **medications not being prescribed,** which meant they were not administered (STF01, STF89).

Assessors

The average score that assessors awarded students for this assessment domain was a 3 (good). The language used by assessors in relation to male students who failed their first attempt (n=6) described students' inability to access stores of self-knowledge. Students were described as having a **haphazard and indecisive approach to patient management** (ACOM36-01, ACOF10-01). Assessors suggested that these students **were lacking in core knowledge** (ASTF39-01, ASTM37-01, ASTF12-01) but did not acknowledge that these limitations could be related to environmental or internal stressors.

Female students who failed their first attempt at the WSE (n=3) were described as **hesitant and lacking in confidence by assessors** (ASTM37-01, ACOM09-02). Female students were observed to **not be systematic in their practice** when they **prescribed medications without examining patients** (ASTF72-01, ACOM08-04) and **did not know the doses of medications** that are prescribed regularly in clinical practice (ACOF22-01, ACOM09-02), which resulted in an **unstructured approach to patient care** (ACOF22-01, ACOM05-01).

Clinical skills

Clinical skills are defined as technical skills (practical procedures), non-technical skills (leadership, teamworking), and cognitive ability (decision-making). (Health Education England (HEE), 2015). In total, three students were assigned to this cluster (2 (m), 1(f)) in the first run of the WSE and 11 students (3 (m),8 (f)) from the second run. Due to the volume of data returned, only the second run of the WSE is reported.

Students

Following the second run of the WSE, STM84 described a lack of confidence in their abilities. STM84 preface all aspects of their performance with the word *tried*: **tried to be polite** (non-technical skills) and **tried to prioritise a sick patient** (cognitive skills). Conversely, STM111 described a systematic approach in their practice: **communicated well with nurses** (non-technical skills), **followed ABC and instigated management** (technical skills).

Following the second run of the WSE, STF14 described becoming increasingly **worried as they managed two patients** that they described as being acutely unwell (technical and cognitive skills). This student focussed all their activity on **assessing and treating these patients** (technical skills). Her assessors noted an improvement in prioritisation but recommended that this student be more **decisive and systematic in their practice** (ACOM16-02, ACOF08-03). STF93 attempted to be logical in their practice by using a **systematic approach** (technical skills) and good teamworking (non-technical skills). The effectiveness of this approach was not observed by the assessors, who described this student as **lacking in structure and having poor teamworking skills.**

Assessors

In relation to the second WSE, there was a convergence between students and assessors regarding their performance. STM84 was described as **acting on impulse rather than being systematic** which led to him to become **increasingly overwhelmed** (ACOF74-01, ACOF46-02). The assessors recognised that STM111 was **systematic in his assessment of patients**, the commencement of the treatment plan and **his use of patient-centred and professional communication** (AGPM46-01, ASTF72-01).

Assessors ACOM16-02 and ACOF08-03, who assessed STF14, observed that this student knew **when to call for senior help** and that they were **faster to assess patients** during the second run of the WSE. The ability to remain focused and calm alluded to STF93. This student's assessors recognised improvements in their **teamworking** (non-technical skills) and **the assessment of the acutely unwell patient** (technical skills), but they generalised that **by this stage, they [the student] should be slicker/quicker** (ACOM36-01, AGPF15-02).

Prioritisation

No students were assigned to this cluster. In total, seven assessors were assigned to this cluster for the first run of the WSE. All male assessors were practising at consultant level within clinical practice (anaesthetics (1), surgical (2) and medical (2)). The three female assessors were practising at the Specialist Trainee level.

Assessors

Statements made in relation to prioritisation could be summarised as the **rapid, effective response** to an untoward event. Those students who passed the WSE were observed as being polite, efficient, **and confident** (ACOM41-01, ACOM27-01, ACOM16-02). In relation to teamworking, they **worked collaboratively** with the nurse and **sought senior help appropriately** (ACOM41-01, ACOM27-01, ACOM16-02). Their practice was deemed to be **systematic when assessing patients** (ACOM41-01, ACOM16-02, ACOM08-04). They developed evidence-based management plans (ACOM27-01, ACOM12-02), and **interventions were**

implemented swiftly and reviewed regularly (ACOM41-01, ACOM27-01, ACOM16-02). Those students who failed to meet this standard were deemed to be **unfocussed, hesitant, indecisive**, and **lacking in basic clinical knowledge** (ACOM41-01, ACOM27-01, ACOM16-02). These standards were unaffected by student gender.

Acutely unwell patient

The context for this cluster was that the nurse would ask the student to come and review an acutely unwell patient as they had concerns about the patient's wellbeing. This normally occurred six minutes into the WSE, and the student would be expected to systematically assess and manage this patient.

Assessors

This domain was only significant in the second run of the WSE. No students were assigned to this cluster. In total, 12 assessors were assigned to this cluster (4(m), 8(f)). All male assessors were practising at consultant level (medical (3), anaesthetics (1)). Female assessors were practising various professional roles within the clinical practice (Consultant (4)–- medical (3), anaesthetics (1); GP (1); ST (3) - medical (2), surgical (1)).

Those students who passed the WSE were observed by male assessors as displaying a **calm, reassuring manner with patients** and **were not distracted by other interruptions** (ACOM36-01, ACOM09-02). They used a **structured approach to assess and examine patients and commenced appropriate treatment plans** (ACOM36-01, ACOM16-02, ACOM09-02). These students reassessed the patient and **recognised the need to seek advice from their senior colleagues** (ACOM36-01, ACOM16-02, ACOM09-02, ACOM08-02). Those who never met the required standard were observed as becoming flustered and **lacking a safe and systematic approach to patient assessment** (ACOM36-01, ACOM16-02, ACOM09-02, ACOM08-02). In one instance, an assessor observed that this resulted in a circumstance where **the student conceded to the decision-making to the nurse!** (ACOM09-02).

Response to interruptions

Students

This cluster was deemed significant for the second run of the WSE and principally assessed how students reacted to the timed interruptions. In total, nine students were assigned to this cluster (55% (m) and 45% (f)). The average score that both male and female students assigned to this domain was a 3 (good).

Male students described that the WSE had given them a better appreciation of the **impact that frequent interruptions have on clinical practice** (STM155). Students described the utilisation of **a job list to minimise the impact of interruptions** on their practice (STM155, STM94).

Only one female medical student (STF161) described their performance as being satisfactory. All her colleagues described a lack of **confidence in their own abilities**, which led to **disorganisation and poor time management** (STF33, STF102, STF163).

Assessors

Regardless of whether a student reviewed their performance positively or negatively, there was convergence with assessors. In total, nine assessors were assigned to this cluster (6 consultants - 2 (m), 4(f); 1 GP -1(f); ST (3) - medical (2), surgical (1)). Responding appropriately to interruptions was observed as being **systematic in the prioritisation of patient care** (ACOF74-01, ACOF46-02, ACOM36-01), **assessing (and re-assessing) patients in a structured manner** (ACOM36-01, ACOM08-02), **instigating appropriate treatment plans** (ACOM36-01, ACOF15-03), **good teamworking and communication skills** (ACOF74-01, AGPF15-02) and **prescribing medications safely and completing all relevant documentation** (ACOF74-01, ACOF46-02, ACOF15-03). No assessors acknowledged the impact that undertaking a second WSE or the relatively short duration of this assessment might have had in affecting the ability of the student to deliver an acceptable level of performance.

Discussion

The Framingham study is considered to be one of the most influential longitudinal studies ever published in the field of cardiology (Oppenheimer, 2010). The Framingham study commenced in 1948 and was the first longitudinal study to examine the epidemiology of cardiovascular disease on 5,209 subjects over a 20-year period (Mahmood, 2014). The full publication of the Framingham study did not occur until 1980 (Dawber, 1980). Due to the nature of longitudinal studies, they are more resource-intensive than single snapshot studies. The volume of data collated over an extended period of time can cause delays in relation to data analysis and the subsequent publication of results, but this does not necessarily mean that the relevance and applicability of this data are lessened (Thomson & Holland, 2003; White & Arzi, 2005).

This article has demonstrated that the process of assessing competence in medical education has been undertaken in similar conditions for thousands of years. This study reported data over a five-year period in relation to one assessment, the WSE. The WSE is still delivered in a format similar to what is described in this publication, so the data on the lived experience of students within this simulated environment could still be considered relevant and valid. This research study identified the intrinsic and extrinsic factors that affected the development of student self-knowledge and the consequences that adhering to a positivist paradigm had on the outcome of a process of assessment, the WSE. Data were collated from the first and second runs of the WSE in 2010, 2012 and 2014. Data analysis of the closed domains in both the students' self-assessment form and the assessors' independent and consensus forms were deemed to be reliable for both the first and second run of the WSE in

2010, 2012 and 2014. The 2nd run of the WSE in 2010 had a lower reported level of reliability than other years and runs. However, it was still acceptable when compared to other assessments used in medical education, such as mini-CEX, DOPS and the OSCE.

Although the assessment tool reported acceptable levels of reliability, it could be contested that the simulated environment lacked the same degree of validity from the perspective of students. The role that students inhabited during the WSE, and the assessment process had an impact on the resultant performance of the students and the level of insight they had into their performance capabilities (which described processing challenges in relation to self-knowledge). This unfamiliarity with the environment, accompanied by the expectation that students would attain a standard of performance that would allow them to progress to graduation, created significant challenges for some students. Students described instances whereby they perceived being in a disempowered state during the WSE, which inhibited the development of momentary and substantial self-knowledge. Croskerry et al. (2013) used the term 'dysrationalia' to describe an inability to think and behave rationally despite adequate intelligence. This term encapsulated the recurring themes identified in students' reflections, where they recounted instances during the WSE where they failed to meet their own performance expectations or those of the assessors. The resultant effect that dysrationalia had on students' performance capabilities (including the development of self-knowledge) was shown to be different for male and female students.

Standpoint theory attests that within any social encounter, there are submissive and dominant parties. In the context of the WSE, there were two prevalent standpoints: that of students (who were perceived to be in a submissive role) and that of the assessors (perceived as being in a dominant role). Instances of convergence between students and assessors were aligned with the objectives of the WSE (to ensure students met the required standard to deliver safe and effective patient care) and the adherence to a shared mental model of the characteristics of male and female medical practitioners. In the first run of the WSE in 2014, convergence between students and assessors was observed in the total number of cases assigned to the 'Communication' and 'Prescribing and Written Documentation' domains. Non-convergence was observed in relation to 'Gender' (both student and assessors) and the domains 'Clinical Skills' and 'Prioritisation'. In the second run of the WSE in 2014, there was convergence in the domains 'Communication', 'Response to Interruptions' and 'Prescribing and Written Documentation' and non-convergence in relation to gender (both student and assessors) and the domains 'Acutely Unwell Patient' and 'Clinical Skills'. Non-convergence was most frequently observed between students and assessors when there was a requirement to undertake a second WSE. In most instances, if a student identified a component of their practice as a strength, the assessors would hold the opposite opinion. The effect of this hierarchical non-convergence could have profound effects on the development of student self-knowledge.

K-means Cluster Analysis identified that a student's gender had a significant influence on the manner in which interventions were delivered during this assessment. Male and female students conformed to a distinct persona that informed the manner in which they practised during the WSE. Male students focussed on addressing those activities that they deemed to be most urgent (for example, assessing the acutely unwell patient) and delegating non-urgent tasks, whilst female students were more holistic in their practice which was characterised as being patient-centred and working collaboratively with the healthcare team. This persona also informed which aspects of a student's performance were given additional attention during the assessment process. These observations are similar to those of Rudland and Mires (2005), who identified that medical students entered a programme of study with a fixed perception of the role of a doctor, and this became more entrenched over time.

The percentage of students whom assessors deemed to have not met the required standard on their first attempt at the WSE was relatively consistent across the three years (2010: (m) 17%, (f) 13%, 2012: (m) 12%, (f) 17%, 2014: (m) 17%, (f) 16%). In all runs of the WSE (bar the second run of the WSE in 2012), the gender of one or both of the assessors was deemed significant. The data provided no clear explanation for this phenomenon as both male and female assessors of all professional roles and specialties were assigned to this domain.

The professional role of an assessor was identified as a determinant of the confidence of an assessor to either pass or fail a student. Female GPs constituted one of the smallest assessor groups (n=4), but they demonstrated the least amount of variation when passing or failing male and female students. Male consultants were the largest assessor group (n=30), and they showed a similar level of agreement to female GPs in their patterns of passing and failing male and female students. Female consultants were more confident failing female students than male students, whilst male GPs were the opposite (they were more confident failing male students than female students). Both male and female STs had almost double the level of variation of female GPs and male consultants in relation to passing and failing male and female students. The judgements made by assessors during the WSE were demonstrated to be statistically significant. So, although there are instances where assessors demonstrated varying degrees of confidence in the decisions they made, it is reasonable to suggest that both gender, professional role and hierarchical deference had limited impact on the outcome of the assessment process. The judgement of assessors was rarely challenged even though changes in the composition of clinical teams had been shown to reduce the amount of time that assessors worked with newly qualified doctors within clinical practice (House of Commons Health Committee, 2008).

Recommendations

Current assessment practices in medical education are time-consuming, resource-intensive, financially prohibitive, vary in reliability, and arguably counterproductive from the

student standpoint (Sood & Singh, 2012). There is a lack of recognition regarding how internal and external factors can affect the assessment process and student performance. The impact of unconscious mental processes, examiner bias and differing perceptions (of a student's gender, ethnicity or an assessor's previous interaction with this person) has been explored with limited depth and rigour. The nature of a student's level of engagement with a simulation activity has been widely discussed within the literature (Dieckmann et al., 2007). The question of an assessor's engagement with the same simulation activity has not been widely explored. The level of immersion and engagement of assessors with simulation activities and their ability to make distinguishable judgements that relate solely to the simulation itself and not clinical practice needs further research.

The WSE adhered to a positivist paradigm and did not consider the concept of self-knowledge in either its design or delivery or in relation to the associated assessment processes. This practice has been shown to be commonplace within medicine and it can result in a clear dissonance between students' experience of undertaking assessments and the perceived need to provide objective measurements as part of a programme of study. Educators need to be cognisant of the impact (both positive and negative) that educational programmes and assessment processes can have on students' ability to develop stores of self-knowledge and to utilise this information within contextual environments (both simulated and clinical).

Further research is required to consider how students might be actively involved in the design and delivery of educational programmes and their associated assessment processes. Enabling students to become co-creators of educational programmes and their associated assessment processes could challenge hierarchical practices and address an expectation for a more transactional approach to education. The legitimisation of students in curriculum design is a new concept within medical education. It is therefore, counter-cultural and disruptive, but arguably a necessary next step in simulation design and delivery. Co-creation could address issues pertaining to student disempowerment (thus minimising disengagement with formal curricula and assessment anxiety) and potentially enhance how simulation activities are delivered in the wider curriculum. These partnerships could also develop conditions within the wider curriculum whereby students develop substantial self-knowledge over several years, thus promoting greater resilience in the individual which would enhance students' ability to access and utilise stores of self-knowledge during assessments. Ultimately, this could allow students to develop self-knowledge appropriate to their stage of professional development.

References

Acton, H. (1951) Comte's positivism and the science of society. *Philosophy, 26*(99), 291–310. http://www.jstor.org/stable/3747190

Aguirre-Raya, K., Castilla-Peón, M., Barajas-Nava, L., Torres-Rodríguez, V., Muñoz-Hernández, O., & Garduño-Espinosa, J. (2016). Self-perception and knowledge of evidence based medicine by physicians. *BMC Medical Education, 16*, 166 -175. https://doi.org/10.1186/s12909-016-0681-6

Alves de Lima, A., Barrero, C., Baratta, S., Castilla Costa, Y., Bortman, G., Carabajales, J., Conde, D., Galli, A., Degrange, G., & Van der Vleuten, C. (2007). Validity, reliability, feasibility and satisfaction of the Mini-Clinical Evaluation Exercise (Mini-CEX) for cardiology residency training. *Medical Teacher, 29*(8). 785-790. https://doi.org/10.1080/01421590701352261

Bleakley, A. (2013). Gender matters in medical education. *Medical Education, 47*, 59-70. https://doi.org/10.1111/j.1365-2923.2012.04351.x

Brannick, M., Erol-Korkmaz, H. T., & Prewett, M. (2011). A systematic review of the reliability of objective structured clinical examination scores. *Medical Education, 45*(12), 1181-1189. https://doi.org/10.1111/j.1365-2923.2011.04075.x

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. http://dx.doi.org/10.1191/1478088706qp063oa

Bruner, J. (1960). *The process of education*. Harvard University Press. Cambridge.

Bynum, W. E., & Artio, A. R. (2018). Who am I, and who do I strive to be? Applying a theory of self-conscious emotions to medical education. *Academic Medicine: Journal of the Association of American Medical Colleges , 93*(6), 874-880. https://doi.org/10.1097/acm.0000000000001970

Cassam, Q. (2014). *Self-knowledge for humans*. Oxford University Press.

Cooney, C. M., Aravind, P., Lifchez, S. D., Hultman, C. S., Weber, R. A., Brooke, S., & Cooney, D. S. (2021). Differences in operative self-assessment between male and female plastic surgery residents: A survey of 8,149 cases. *American Journal of Surgery, 221*(4), 799–803. https://doi.org/10.1016/j.amjsurg.2020.04.009

Cordovani, L., Tran, C., Wong, A., Jack, S. M., & Monteiro, S. (2023). Undergraduate learners' receptiveness to feedback in medical schools: A scoping review. *Medical Science Educator, 33*(5), 1253–1269. https://doi.org/10.1007/s40670-023-01858-0

Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 1: Origins of bias and theory of debiasing. *British Medical Journal, Quality and Safety, 22*(Supple 2), 58–64. https://doi.org/10.1136/bmjqs-2012-001712

Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: A systematic review. *Journal of the American Medical Association, 296*(9), 1094–1102. https://doi.org/10.1001/jama.296.9.1094

Dawber, T. (1980). *The Framingham study: The epidemiology of atherosclerotic disease*. Harvard University Press. https://

doi.org/10.4159/harvard.9780674492097

Dias Pereira, M. A., & Barbosa, M. A. (2013). Teaching strategies for coping with stress – the perceptions of medical students. *BMC Medical Education, 13*, 50. https://doi.org/10.1186/1472-6920-13-50

Dieckmann, P., Gaba, D., & Rall, M. (2007). Deepening the theoretical foundations of patient simulation as social practice. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare, 2*(3), 183-193. https://doi.org/10.1097/SIH.0b013e3180f637f5

Domician, M., & Éva, D. (2017). Measuring the accuracy of self-assessment among undergraduate students in higher education to enhance competitiveness. *Journal of Competitiveness. 9*(2), 78-92. http://dx.doi.org/10.7441/joc.2017.02.06

Eggleton, K., Goodyear-Smith, F., Paton, L., Falloon, K., Wong, C., Lack, L., Kennelly, J., Fishman, T., & Moyes, S. (2016). Reliability of Mini-CEX assessment of medical students in general practice clinical attachments. *Family Medicine, 48*(8), 624-630.

Elliott, V. S., Dekhtyar, M., Pouwels, M. V., & Skochelak, S. E. (2019). The American Medical Association medical education innovation challenge: Incorporating the student voice into medical education transformation. *Academic Medicine: Journal of the Association of American Medical Colleges, 94*(9), 1343–1346.

Epstein, R. (2007). Assessment in medical education. *The New England Journal of Medicine, 356,* 387-396. https://doi.org/10.1056/NEJMra054784

Eva, K. W., & Regehr, G. (2011). Exploring the divergence between self-assessment and self-monitoring. *Advances in Health Sciences Education: Theory and Practice, 16*(3), 311–329. https://doi.org/10.1007/s10459-010-9263-2

Gardiner, F. (2016). The art of self-knowledge and deduction in clinical practice. *Annals of Medicine and Surgery, 10,* 19-21. https://doi.org/10.1016/j.amsu.2016.07.008.

General Medical Council (GMC). (2015). *Promoting excellence: Standards for medical education and training.* General Medical Council. London.

General Medical Council (GMC). (2018). *Outcomes for graduates.* General Medical Council. London.

General Medical Council (GMC). (2019). *Good medical practice.* General Medical Council. London.

Govern, J. M., & Marsch, L. A. (2001). Development and validation of the situational self-awareness scale. *Consciousness and Cognition, 10*(3), 366-378. https://doi.org/10.1006/ccog.2001.0506

House of Commons Health Committee (2008). *Modernising medical careers: Third Report of Session 2007-08.* The Stationery Office Limited. London.

Harden, R. M., & Gleeson, F. A. (1975). Assessment of clinical competence using an Objective Structured Clinical Examination (OSCE). *Medical Education, 13*(1), 39-54.

Health Education England (East of England) (HEE). (2015). *Simulation strategy 2015.* Health Education England. Leeds.

Jovic, E., Wallace, J. E., & Lemaire, J. (2006). The generation and gender shifts in medicine: An exploratory survey of internal medicine physicians. *BMC Health Services Research, 6*(55), 1-10. https://doi.org/10.1186/1472-6963-6-55

Ker, J. S., Hesketh, E. A., Anderson, F., & Johnston, D. A. (2006). Can a ward simulation exercise achieve the realism that reflects the complexity of everyday practice junior doctors encounter? *Medical Teacher, 28*(4), 330-334. https://doi.org/10.1080/01421590600627623

Kokushkin, M. (2014). Standpoint theory is dead, long live standpoint theory! Why standpoint thinking should be embraced by scholars who do not identify as feminists? *Journal of Arts and Humanities, 3*(7), 1-20.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Kukulski, P., Ahn, J., Babcock, C., Cheema, N., Clayton, G. C., Olson, A. S., Olson, N., Tekwani, K. L., & Carter, K. (2021). Medical student self-assessment as emergency medicine residency applicants. *AEM Education and Training, 5*(3), 1–7. https://doi.org/10.1002/aet2.10578

Long, S., Rodriguez, C., St Onge, C., Tellier, P., Torabi, N., & Young, M. (2022). Factors affecting perceived credibility of assessment in medical education: A scoping review. *Advances in Health Sciences Education, 27*, 229–262. https://doi.org/10.1007/s10459-021-10071-w

Mahmood, S. S., Levy, D., Vasan, R. S., & Wang, T. J. (2014). The Framingham heart study and the epidemiology of cardiovascular diseases: A historical perspective. *Lancet, 383*(9921), 999–1008. https://doi.org/10.1016/S0140-6736(13)61752-3

McGaghie, W. C. (2015). Mastery learning: It is time for medical education to join the 21st century. *Academic Medicine: Journal of the Association of American Medical Colleges, 90*(11), 1438-1441. https://doi.org/10.1097/ACM.0000000000000911

McIlwaine, L. M., McAleer, J. P. G., & Ker, J. S. (2007). Assessment of final year medical students in a simulated ward: Developing content validity for an assessment instrument. *International Journal of Clinical Skills, 1*(1), 33-35.

Oppenheimer, G. (1980). Framingham heart study: The first 20 years. *Progress in Cardiovascular Diseases, 53,* 55–61. https://doi.org/10.1016/j.pcad.2010.03.003

Owens, H. (2012). Early use of simulation in medical education. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare, 7*(2), 102–116. https://doi.org/10.1097/SIH.0b013e3182415a91

Rees, K. L. (2013). The role of reflective practices in enabling final year nursing students to respond to the distressing emotional challenges of nursing work. *Nurse Education in Practice, 13*(1), 48-52. https://doi.org/10.1016/j.nepr.2012.07.003

Rudland, J. R., & Mires, G. J. (2005). Characteristics of doctors and nurses as perceived by students entering medical school: Implications for shared teaching. *Medical Education, 39*(5), 448-455. https://doi.org/10.1111/j.1365-2929.2005.02108.x

Schumann, C. (2016). Knowledge for a common world? On the place of feminist epistemology in philosophy of education. *Education Sciences, 6*(1), 1-13. https://doi.org/10.3390/educsci6010010

Sharma, M. (2019). Applying feminist theory to medical education. *The Lancet, 393*(10171), 570-578. https://doi.org/10.1016/S0140-6736(18)32595-9

Sood, R., & Singh, T. (2012). Assessment in medical education: Evolving perspectives and contemporary trends. *The National Medical Journal of India, 25*(6), 357–364.

Stirling, K., Smith, G., & Hogg, G. (2012). The benefits of a ward simulation exercise as a learning experience. *British Journal of Nursing, 21*(2), 116-122. https://doi.org/10.12968/bjon.2012.21.2.116

Tejinder, S. A. (2022). *Principles of assessment in medical education* (2nd ed). Jaypee Brothers Medical Publishers.

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation, 27*(2), 237-246. https://doi.org/10.1177/1098214005283748

Thomson, R., & Holland, J. (2003). Hindsight, foresight and insight: The challenges of longitudinal qualitative research. *International Journal of Social Research Methodology, 6*(3), 233–244. http://dx.doi.org/10.1080/1364557032000091833

Till, H., Ker, J., Myford, C., Stirling, K., & Mires, G. (2015). Constructing and evaluating a validity argument for the final-year ward simulation exercise. *Advances in Health Sciences Education, 20*(5), 1263-1289. https://doi.org/10.1007/s10459-015-9601-5

Tsui, K., Liu, C., Lui, J., Lee, S., Tan, R., & Chang, P. (2013). Direct observation of procedural skills to improve validity of students' measurement of prostate volume in predicting treatment outcomes. *Urological Science, 24*(3), 84 - 88. https://doi.org/10.1016/j.urols.2012.09.001

Watson, M. J., Wong, D. M., Kluger, R., Chuan, A., Herrick, M. D., Ng, I., Castanelli, D. J., Lin, L., Lansdown, A. K., & Barrington, M. J. (2014). Psychometric evaluation of a direct observation of procedural skills assessment tool for ultrasound-guided regional anaesthesia. *Anaesthesia, 69*, 604–612. https://doi.org/10.1111/anae.12625

White, R., & Arzi, H. (2005). Longitudinal studies: Designs, validity, practicality, and value. *Research in Science Education, 35*, 137–149. https://doi.org/10.1007/s11165-004-3437-y

Williams, B., Beovich, B., Ross, L., Wright, C., & Ilic, D. (2017). Self-efficacy perceptions of interprofessional education and practice in undergraduate healthcare students. *Journal of Interprofessional Care, 31*(3), 335-341. https://doi.org/10.1080/13561820.2017.1286637

Copyright: © 2024. Kevin Stirling, Angela Roger and Keith Topping. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.