British Journal of Medicine & Medical Research
3(4): 1806-1817, 2013

SCIENCEDOMAIN *international*
www.sciencedomain.org

An Alternative Certification Examination 'ACE': Can Post Graduate Methods Be Used to Assess Clinical Skills in Medical Under Graduates

M. Morris[1*], M. Bennett[2], M. Hennessy[1], K. C. Conlon[1,3] and P. F. Ridgway[1,3]

[1]Education Division, School of Medicine, University of Dublin, Trinity College. Dublin 2, Ireland.
[2]Emergency Department, Tallaght Hospital, Dublin 24, Ireland.
[3]Department of Surgery, University of Dublin, Trinity College, Tallaght Hospital Campus, Dublin 24, Ireland.

Authors' contributions

This work was carried out in collaboration between all authors. Authors MM, MB, MH, KCC and PFR designed the study, authors MM, MB and PFR performed the statistical analysis. Authors MM, MB, MH, KCC and PFR wrote the protocol, and authors MM, MB and PFR wrote the first draft of the manuscript. Author MM, MB and PFR managed the analyses of the study. Authors MM, MB and PFR managed the literature searches. Authors MH, KCC and PFR revised the first draft. All authors read and approved the final manuscript.

Research Article

Received 25th March 2013
Accepted 27th May 2013
Published 16th June 2013

ABSTRACT

Background: Progressing from undergraduate education to post graduate training has been punctuated by a clinical examination which has not changed significantly in decades. This study investigated the feasibility of using a validated postgraduate assessment methodology in an undergraduate setting; The Toronto Patient Assessment & Management Exam (PAME).
Methods: A standardised patient-centred multifaceted healthcare pathway examination consisting of 4 separate consecutive encounters was piloted in the final year of undergraduate training. The entire final year medical class was invited to participate. The final sample of 25 was selected on a consecutive, volunteer basis. Student's experienced 2 standardised simulated cases; 1 medical, 1 surgical. Candidates were examined by 2 independent examiners (subject experts) and were ranked on a Global

**Corresponding author: Email: morrism4@tcd.ie;*

Rating Scale. Passing standard was set at 3/5 - 'barely adequate for Intern/PGY1' but with the addition of second pass criteria of avoidance of an egregious error.
Results: 23 students completed the examination. Two arrived late and were excluded. 21/23 demonstrated knowledge and skills at least at minimum expected standard. 18/23 avoided an egregious error. Subgroup analysis identified better performance in the assessment and management of the medical case and the review encounter (encounter 4) was the lowest scoring in both cases. The format was well received by students and examiners.
Conclusion: The use of an alternative certification examination 'ACE' based on a postgraduate format 'PAME' in undergraduate setting appears feasible and discriminatory. Inclusion in the pass criteria of avoidance of egregious error appears to improve the specificity of the examination. The ACE format reveals potential to replace elements of prepractice (PGY1) clinical barrier assessment.

Keywords: Assessment; competence; validity; safety to practice; medical graduates.

1. INTRODUCTION

Ensuring specificity in a certification examination in medicine remains a challenge to Educators. Consequently we experience false positives and poor students pass. Historically assessment in a primary medical degree programme focused on measuring a knowledge base. Although these tests of knowledge are unquestioningly important, they are also incomplete appraisals if we believe there is more to the practice of medicine than knowing. Graduates of medical training must also be able to 'show how' [1] to use the knowledge they have accumulated. Acquirement of competence in core clinical skills requires a more holistic approach. Students must attain information from a variety of sources, assimilate these findings and ultimately translate such findings into a rational diagnosis with a resultant instigation of a management plan. A further challenge to the objective assessment of competence is with regard to the validity of current practices. Here examinations in artificial settings are accepted as accurately predictive of what a graduate might actually do when functioning with relative independence in a clinical setting.

Undergraduate assessment is a prerequisite to licensure, the basic qualification to practice in a health profession. Currently the 'long case' is the most common method of certification assessment in a primary medical degree programme. The long case consists of a student spending 40 minutes alone with a patient, interviewing the patient, examining the patient, reviewing medical charts and presenting their findings to 2-3 Senior Consultants. Although popular due to the authenticity of the examination being based in a Hospital setting with In-patients, many limitations have been reported. These include issues with reliability, validity, lack of direct observation, only narrowed elements of a wide curriculum assessed [2,3,4,5] and poor case specificity [6] (Table 1). Extrapolation of demonstrated competence in one or two isolated cases to a broader core curriculum warrants caution. Increasing the number of long cases has been suggested as a means to address these deficits [7]; however, the feasibility of this is questionable given the demand on finite resources of Senior Clinical Staff as Examiners and sufficient patients clinically well enough to participate.

This paper seeks to investigate whether the previously validated post graduate 'PAME' methodology [8] (Table 2) could be utilized in an undergraduate setting to assess medical and surgical cases at the point of certification. The 'PAME' assessment method involves observed encounters at multiple time points with the same patient resulting in the students

being required to manage an entire medical and surgical episode. Whether this more integrative approach will identify student's preparedness for the intern /PGY1 role will be explored.

Table 1. Long case- strengths and limitations

Assessment type	Positives	Negatives
Long cases	• History and examination • Authentic – real patients	• Not directly observed • Poor generalisability • Low reliability • Inability to sample the curriculum widely • Practical skills not examined

Table 2. Patient assessment and management examination 'PAME'

Purpose: The Patient Assessment and Management Examination (PAME) was developed to assess the clinical competence of surgery residents. ***Method****: S*enior residents participated in a six-station PAME that used standardized patients (SPs). PAME stations were 30 minutes each and designed to simulate a more comprehensive approach to patient encounters. Surgery faculty completed all rating forms. Station Component **1. Conduct initial patient assessment**: Referral letters and test results were supplied and the resident was expected to complete a history and physical examination, and obtain informed consent if needed. **2**. **Order and interpret tests** **3. Conduct a second encounter with patient**: The resident was expected to discuss diagnosis and develop a management plan. **4. Respond to an oral examination pertaining to the case**

2. METHODS

2.1 Setting

This study occurred in a large teaching Hospital affiliated to a Dublin University Medical School.

2.2 Research Sample

A sub-group (n= 26) of a final year medical class were allocated by the Medical School to the Teaching Hospital for clinical skills training. All were invited to participate. One student chose not to participate. The final 25 were selected on a consecutive, volunteered basis. Two participants arrived after the exam had commenced and so were not permitted to participate thus the final sample n=23.

2.3 Exam Format

All students assessed 2 standardized simulated patients one presenting with a common surgical complaint (right iliac fossa pain) and a second with a common medical complaint (central chest pain) across 4 different encounters as follows - Table 3.

Table 3. Examination process

Encounter 1: Focused history and examination	12 minutes
Encounter 2: Review of investigations	5 minutes
Encounter 3: Diagnosis management plan	6 minutes
Encounter 4: Review examination	6 minutes
	Total exam time = 29 minutes
	1 minute to move rooms from Medical to surgical case

The participant entered an exam centre and remained in that centre with the patient and 2 Examiners for the entire 29 minutes. Participants were examined by 2 Surgeons in the Surgical case and 2 Physicians in the Medical case. A bell rang after each of the encounters allocated times had passed to move participants onto the next encounter.

2.4 Pass Criteria/Egregious Error

Safety for certification in core skills was assessed by the ability of the participant to score at least 3/5 (barely adequate for PGY1 practice) in all four Encounters and avoidance of an egregious error. The rationale for the inclusion of an egregious error was to ensure those achieving a pass mark> 3 were indeed safe to practice. Irrespective of the students allocated scores across the 4 encounters - failure to avoid an egregious error in each case resulted in an overall score of 2 i.e. Fail. The egregious errors included ruling out a potential ectopic pregnancy in the female patient with right iliac fossa pain and demonstrating the ability to differentiate between an ST elevation Myocardial Infarction (MI) and a non ST elevation MI in the patient with chest pain with regard to the appropriate use of thrombolytics.

2.5 Validity of Tools

Previously validated global assessment tools from the 'PAME' [8] methodology were slightly modified. To maintain validity modifications were minimal and consisted of only the assessment category terminology being amended to reflect a level appropriate to Intern/PGY1 level of competence. (Appendix 1).

2.6 Reliability of Data

All examiners were furnished with the standardized marking sheets and exam format prior to the examination. Standard instructions were issued to all examiners prior to the examination and again on the morning of the Examination. Pass/fail criteria was discussed including egregious errors. Two examiners who were subject experts assessed each candidate at

each case across all 4 encounters and marked independent of each other. Surgeons assessed the surgical case and Physicians assessed the medical case.

2.7 Assessment Criteria

Candidates were marked using a global rater. A mark of at least 3 (range 1-5) was deemed the minimum standard for safety to practice supervised as an Intern/PGY1. Candidates were assessed across 4 encounters with each patient -which involved eliciting a focused history, performing a physical examination, appropriate communication - information giving, gathering, analysis and inter-personal skills. Time and overall case management for both the medical and surgical cases were assessed. Examiners finally awarded a global score based on overall performance. Students making an egregious error would have a red sticker applied to their marked sheets and be scored a 2/5 and thus fail that case.

2.8 Feedback to Participants

All participants received a written copy of their marking sheets with Examiners comments and were offered the opportunity to meet and discuss their performance with the study coordinators within 1 week of the examination.

2.9 Evaluation of the Alternative Exam Format

All participants and examiners completed an evaluation of the exam format and experience.

2.10 Data Analysis

Non parametric tests [9] were undertaken to identify median scores in each encounter and across both the medical and surgical cases. A Wilcoxin signed rank test was performed to investigate sub-group analysis between the two cases [9].

3. RESULTS

3.1 Overall Performance

The pass criterion was that participants should perform at least at ‘a barely adequate’ level (Score 3/5) for internship/ PGY1 in all subsections of history, examination, communication, time and case management and avoid an egregious error. 21/23 of students participating achieved this standard. 2 performed less than this standard scoring 1 or 2/5 across the 4 encounters in both the medical and surgical cases. 18/23 avoided an egregious error thus a total of 5 made an egregious error - 2 in the medical cases and 3 in the surgical case. 3 students met the pass criterion and in the absence of the inclusion of avoidance of egregious error would have been deemed at a level adequate for Internship/PGY 1 practice. The 2 students who performed below the accepted standard both made an egregious error in both the medical and surgical cases. There was no difference in median scores between the medical and surgical case overall (4 encounters together as a total) with a median score of 4 (Q1:3, Q3:4) being demonstrated in both. Each individual encounter score is reported in Table 4.

Table 4. Median Scores over all 4 Encounters

Format	Median score (Q1:Q3) medicine	Median score (Q1:Q3) surgery
Encounter 1 : History	4(3.5:5)	4(3:4)
: Examination	4(3:4)	4(3:4)
: Knowledge	4(4:4.5)	4(4:4)
: Communication	4(4:4)	4(4:4)
:Time Management	4(4:4)	4(4:4)
Encounter 2 : Use of Resources'	4(4:5)	4(4:4)
: Interpretation of data	4(4:4)	4(3:4)
Encounter 3: Management of case	4(4:5.5)	4(4:4)
:Knowledge of case	4(4:5)	4(3.5:4)
:Communication	4(4:5)	4(4:4)
:Response to patient concerns	4(4:4)	4(4:4)
Encounter 4 : Management of problem	4(4:4)	4(3:4)
:Knowledge of problem	4(4:5)	4(3:4)

3.2 Sub-group Analysis-Comparison of Scores between Medicine and Surgery Case

In Encounter 1 participants performed better in history taking in the medical case compared to the surgical case - this difference was statistically significant with a p value of 0.00015.- Table 5. There was no significant difference between cases for participants physical examination abilities on standardized patients p= 0.413. With regards to knowledge of the case participants demonstrated greater knowledge of the medical case -chest pain and myocardial infarction that they did the surgical case-abdominal pain and appendicitis and appendectomy procedure. This difference was not statistically significant with a p value = 0.0243. Regarding overall communication with the patient in Encounter 1, participants performed better in the medical case - Examiners noted better inter-personal and listening skills in the medical case when compared to the surgical Examiners assessments. This difference was statistically significantly with a p value of 0.008 Table 5. With regard to Time Management there was no difference in skills demonstrated between cases- p= 1.

In Encounter 2, 'examining appropriate use of resources', participants scored better in the medical case that the surgical case. This difference was statistically significant with a p value = 0.0007. There was no differences demonstrated between cases with regard to interpretation of blood results, ECG's and radiography p= 0.142. In Encounter 3 – 'Case Management', there was little difference between the cases with regard to 'management plan' (p = 0.14), 'knowledge' (p= 0.68) and 'Response to patient concerns' (p= 0.10). However with regard to 'communication' Examiners reported that participants communicated better with patients in the medical case providing superior explanations of diagnosis and management of the presenting complaint compared to the surgical case. This difference between cases was statistically significant with p = 0.0138 Table 5.

In Encounter 4, 'Review', there was little difference between the cases with regard to 'management of the complication' p = 0.728, 'knowledge' p= 0.84 and 'Response to patient concerns' p= 0.22. However with regard to communication Examiners reported that students communicated better with the patient in the medical case with regard to explanations of

complications and management of same. This difference between cases was statistically significant with p = 0.0267 Table 5.

Table 5. Wilcoxin - sub-group analysis - performance on medical vs. surgical case

Format	p value	Superior performance
Encounter 1 : History	0.00015	Medicine case
: Examination	0.4137	
: Knowledge	0.0243	Medicine case
: Communication	0.008	Medicine case
: Time Management	1	
Encounter 2 : Use of Resources'	0.0007	Medicine case
: Interpretation of data	0.142	
Encounter 3: Management of case	0.1475	
: Knowledge of case	0.682	
: Communication	0.0138	Medicine case
: Response to patient concerns	0.101	
Encounter 4 : Management of problem	0.728	
: Knowledge of problem	0.8464	
: Communication of plan to patient	0.0267	Medicine case
: Response to patient concern s	0.225	

3.3 Qualitative Results

3.3.1 Participant evaluations

All participants completed a LIKERT scale to evaluate the 'ACE' format and had the opportunity for free text on the form. The responses for each statement were generally positive. All participants reported that the 'ACE' format was a useful learning experience and that they preferred to be examined by 2 Subject Experts rather than 1. Participants reported that the time lines of 30 minutes per case proved adequate for the surgical case but proved insufficient for the medical case in this examination format. On discussion it was revealed that many medical patients have co-existing diseases thus have a longer more complicated medical and drug history which required more time than allocated in the 'ACE' format. Participants found 30 minutes sufficient for all 4 encounters with the surgical case. Participants reported that they would prefer unfamiliar simulated patients to add to the authenticity of the examination.

Participants reported very positive attitudes to this new method of assessment specifically related to the integration of previously acquired knowledge and skills. Free text comments from participants included the following, 'this exam looks at the bigger picture of managing a patient episode, not just doing a history or doing a cannuale'. 'I think it's a more suitable exam to determine who will be a safe doctor but the time frames are too regimented'. 'Not having real patients is a drawback -having to palpate the abdo when the patient wasn't in pain was difficult'. 'Good format as it's helped me direct my study'. 'Good as the format allowed you to formulate your own management plan including follow-up'. 'Too short - needed more time. 'Needs real life patients then would be great'. 'I like that this format allowed you to follow the patient through their Hospital stay but the section for review was a bit much as this wouldn't be an Intern role'. 'Definitely has helped me direct my studies - need more clinical information - big stuff!'.

3.3.2 Examiners evaluations

All Examiners completed a LIKERT scale to evaluate the 'ACE' examination process and had the opportunity for free text on the form. The responses to each statement were generally positive. All Examiners reported that the 'ACE' format was superior to the OSCE in preparing students for managing a patient in the clinical setting. Examiners also reported that the time lines of 30 minutes per case proved adequate for the surgical case but proved insufficient for the medical case in this examination format as medical cases are often more complex. Examiners agreed with participants that unfamiliar simulated patients were required in future examinations.

Examiners were positive regarding the exam format. They agreed that the integrative approach with case management clearly differentiated between good and weak students reporting – 'there was nowhere to hide'. Examiners of the surgical case reported that the 'ACE' format clearly highlighted deficits in knowledge in history taking and examination and they welcomed the addition of avoidance of egregious errors as passing criteria on safety to practice grounds. Eliciting a gynaecological history was identified as an area of weakness which would be subsequently addressed in the surgical teaching programme. Examiners reported that if these limitations were addressed the 'ACE' format should be evaluated against the traditional long case.

4. DISCUSSION

A realistic, holistic, reliable, valid and integrated certification examination remains elusive despite great efforts by many [2,3,4,5]. An examination format that reliably measures the breadth of knowledge and skill acquired over the entirety of the medical curriculum is paramount to ensure only safe students pass. This also ensures credibility of the assessment process meeting faculty and societal expectations.

The inclusion of avoidance of egregious errors as pass criteria in a 'ACE' format resulted in more failures - 3 students would have passed a traditional examination on accrued scores yet they failed to consider a potentially life threatening condition in the patient. This has implications on safety to practice yet is not currently an assessment criterion in certification examinations in Medicine. It could be argued that inclusion of egregious errors in the standard setting programme may increase specificity -this warrants further investigation with larger numbers. Participants performed well in the first 3 encounters with the lowest scores in the forth 'review' encounter. This is not unexpected as this level of intervention would traditionally be at Senior House Officer/PGY2 level. This finding indicates that the PAME [8] format can be extrapolated to a more junior pre certification level provided the complexity of the cases utilized are appropriate to the expected level of clinical functioning. Students performing poorly did so across all 4 encounters as well as making an egregious error validating the exam format. The marked differences in knowledge and skill in managing the medical and surgical case warrants further investigation and may be a reflection of current teaching methods. Student's demonstrated less knowledge of surgical procedures and potential complications compared to medical problems. Eliciting a gynaecological history was also an area of weakness identified in the surgical case.

The reason cited for a lower score in Encounter 2 by Surgical Examiners was due to participants suggesting a CT scan as a first line investigation prior to blood profiles and radiography. The opting for CT scanning as a first line investigation may be a reflection of a

cultural difference in the utilization of high end technology for diagnosis as many students originate from North American or Canada.

There was no statistically significant difference in the student's results with regard to time management between cases yet both Examiners and Participants reported that 30 minutes was insufficient time for the medical case. This was reported as being due to patients with medical problems having more co-existing diseases requiring a longer interview to elicit a full past medical and drug history. Results indicate that students and examiners found the 'ACE' format comprehensive and discriminatory by identifying those not yet at a level expected to provide safe patient care. This exam format also has potential as a practice/mock exam to remediate those performing below Intern/PGY1 level ensuring all graduates of medical degrees are meeting regulatory standards for entry into the work force [10].

4.1 Limitations

Within this project there is a potential bias as participants self selected to partake however their allocation to the Teaching Hospital occurred centrally. The Exam co-ordinators have a close working relationship with participants which may have influenced their decision to participate. The timeframes may be too restrictive potentially limiting the participant's performance. Familiar standardized simulated patients were utilized which may have negatively impacted on the authenticity of the examination. Participants having the opportunity for formative assessment with written and verbal feedback may explain the positivity towards the examination rather than the 'ACE' format itself. Participants only experienced 2 cases so extrapolation to overall ability is limited.

5. CONCLUSION

In conclusion**,** ensuring medical graduates are safe to practice is mandatory for all training institutions [10]. It is essential that the final certification examination process has good sensitivity and more importantly specificity to ensure only safe students pass. The current gold standard certification examination in Medicine (Long Case) has well documented limitations [2,3,4,5,6]. This alternative certification examination based on the 'PAME' format appears feasible in a primary degree setting. Its integrative process may assist with the transferability of skills to the clinical setting.

Inclusion of avoidance of an egregious error may improve the specificity of the assessment. Research including psychometrics is indicated to accurately quantify how many cases are required in a certification examination to sample the curriculum widely and eliminate false positives. Ultimately, students need to be assessed on multiple cases on multiple occasions using multiple assessors [7] which remains an ongoing challenge to Educators.

CONSENT

N/A - Actors utilized as patients.

ETHICAL APPROVAL

Ethical approval was given by the Faculty of Health Sciences Ethics Committee, The University of Dublin, Trinity College.

COMPETING INTERESTS

The authors report no competing interests or conflict of interest in researching and writing this paper.

REFERENCES

1. Miller GE. The assessment of clinical skills /competence/performance. Acad Med. 1990;65(9)(Suppl):63-7.
2. Van der Vleuten C. Validity of final examinations in undergraduate medical training. BMJ. 2000;321(7270):1217-9.
3. Wass, V, Jolly B. Research paper: does observation add to the validity of the long case? Med Educ. 2001;35:729-34.
4. Wass, V, van der Vleuten C. The long case. Med Educ. 2004;38:1176-80.
5. Norcini J. The validity of the long case. Med Educ. 2001;35:720-1.
6. Eva KW. On the generality of specificity. Med Educ. 2003;37(7):587-8.
7. Ponnamperuma GG, Karunathilake IM, McAleer S, Davis MH. The longcase and its modifications: a literature review. Med Educ. 2009;43:936-941.
8. MacRae HM, Cohen R, Regehr G, Reznick R, Burstein M. A new assessment tool: the patient assessment and management examination. Surg. 1997;122:335-44.
9. Greene J. Learning to use statistical tests in psychology. Second Edition Open University Press. Bucks. UK; 1999.
10. Irish Medical Council. The Medical Practitioners Act., Dublin. Ireland; 2007.

APPENDIX

Appendix 1. Pame Summary

Toronto General Hospital

Exam Procedure

This examination is made up of 2 tracks of 6 stations (each 25 min in length). The tracks are identified by colour as track **Yellow (exam rooms)** or track **Blue (conference rooms)**. Each resident completes all 6 stations on a single track.

Each examiner and resident ID badge is colour coded to track.

Each station has one Examiner and one Standardized Patient.

The procedure at each station is as follows:

Bell – End of Previous Station

2 Minutes

Examinee move to next station and reads on door, **referral letter and first encounter information.** Examinee gives ID stickers to examiner (will need 5 total - one for each GRS page)

Bell – Candidates Enter Room

I) First Encounter (laminated paper) - 7 Minutes

Examinees perform focused history and physical examination on standardized patient (SP).

Whistle - SP leaves room.

II) Investigations - 3 Minutes

Examiner asks what investigations candidate would order (may write on investigations paper provided) and provides candidate with results.

All results are in folder on desk at each station. X-ray images are also provided at some stations on the laptop. (Specific details at each station)

Whistle - SP returns to room

III) Second Encounter (laminated paper that the Examiner hands to Examinee) – 8 Minutes

The examinee then gives SP diagnosis/treatment plan, answers SP's questions, and obtains informed consent.

N. B. Verbal consent will be obtained instead of signing consent form for time/logistical reasons.

***Whistle* - SP leaves room**

IV) Structured Oral Examination – 5 Minutes

Examiner then asks pre-determined questions provided that is related to the case and fills in the Global Rating scale forms.

Add both examinee and examiner stickers – **one to each page!!!**

Bell – End of Station

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=205&id=12&aid=1508