

Institution: The University of Leeds		
Unit of Assessment: 23 Education		
Title of case study: Improving the quality and efficiency of healthcare professional education and training via enhanced assessment practices		
Period when the underpinning research was undertaken: 2006-2019		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Matt Homer	Associate Professor	2002 -
Richard Fuller	Professor	2005-2018
Godfrey Pell	Principal Research Fellow	2002-2018
Period when the claimed impact occurred: August 2013 - 2020		
Is this case study continued from a case study submitted in 2014? No		
1. Summary of the impact (indicative maximum 100 words)		
<p>For patients to receive high-quality care, medical schools and national healthcare regulators must ensure that only competent trainees are allowed to practise, but effective healthcare assessment is a highly complex, high stakes and costly process. Research by Homer and colleagues has enhanced the assessment practice, design of assessment instruments and psychometric quality-assurance approaches in medical schools and national healthcare examinations nationally and internationally. Their work has enabled the General Medical Council (GMC) and other national healthcare regulators to improve their assessment practices in five nationally important licensing exams, both in the UK and overseas. This work improves the quality of a range of healthcare practitioners entering and progressing in the workforce, saves on resources, and in the long-term enhances patient safety.</p>		
2. Underpinning research (indicative maximum 500 words)		
<p>The Leeds research has focussed on evidencing the benefits of superior quality assurance processes and assessment designs in costly, complex and high stakes settings. This empirical and theoretical work has underpinned the development of assessment research in medical education and other disciplines (e.g. dentistry, pharmacy, social work, and nursing). Objective Structured Clinical Examinations (OSCEs) are complex, summative assessments of clinical, practical and communication skills where candidates (often 250+) rotate around a set of clinical cases (typically 15 or so), and engage as junior doctors with real or simulated patients (trained actors). Clinically trained assessors rate candidate performance – both holistically in a global measure of performance for the case, and in a more fine-grained way via a ‘key features’ checklist. OSCE-type assessments - also used in other healthcare professions such as pharmacy, midwifery and nursing – are seen as the ‘gold standard’ for assessing professional performance, but are resource intensive, costly, prone to error, and require strong evidence to motivate substantial institutional change in practice and design.</p>		
Identifying the resit problem with OSCEs		
<p>Fifteen years ago a group of researchers at the Schools of Education and Medicine at the University of Leeds began a research programme to monitor and improve the quality of our own OSCE, and other assessments on the Leeds undergraduate medical degree. Early research was partly motivated by the problem of repeatedly failing individual students, who passed subsequent resits (both OSCE and applied knowledge tests), but who did not show sustained longitudinal improvement in their performance [1]. This research demonstrated that these assessment practices were not fit for purpose and passed some students when they should not. Over time, such under-performing students might end up as junior doctors who struggle in practice, a serious and costly problem for patients, medical managers and taxpayers. This ‘resit problem’ remains a problem across medical education, and there are similar issues in performance assessments across a range of healthcare professions, including dentistry, nursing and pharmacy, and in other professions (social work, law).</p>		

Developing sequential testing as a solution to the resit problem

Having identified the resit problem [1], the group began researching the possibility of improving models of assessment design. This work led to developing and implementing sequential models of assessment for OSCEs, where students who perform sufficiently well are only required to do the first part of a two-part sequence, with only the weakest students taking both parts [3,5]. The research on sequential testing shows that these assessment models produce better 'diagnostic' decisions in the trainees who should and should not progress and solve the 'resit' problem identified earlier, with fewer students getting through exams but not sustaining their learning progress over time [5]. Because OSCE-type examinations are expensive to deliver, and managing weak students is costly in academic resources, the research [3] evidences significant cost savings.

Assessor issues and improving quality in OSCEs

In standard OSCE designs, candidates are examined by different sets of assessors, who can often vary in the stringency of their marking. The research [2,4] developed analytical processes that use OSCE assessment data to identify such variations and make adjustments to improve the assessment quality. The innovative approach is based on 'assessing the assessment' via a range of case- and exam-level metrics (i.e. psychometric indicators) which feed back into assessment design and improve quality over time. The research [1,2,4,6] focussed on creating a set of robust *post hoc* analysis methods for assessing clinical performance, particularly measuring the 'assessor' effect, and broader ways to judge whether each clinical case, and the exam as a whole, are 'working' well psychometrically in the assessment outcomes they produce.

3. References to the research (indicative maximum of six references)

- [1] Pell G., Fuller R., Homer M., Roberts T. 2012. Is short-term remediation after OSCE failure sustained? A retrospective analysis of the longitudinal attainment of underperforming students in OSCE assessments. *Medical Teacher* 34 (2):146–50. <https://doi.org/10.3109/0142159X.2012.643262>.
- [2] Pell G., Fuller R., Homer M., Roberts T. 2010. How to measure the quality of the OSCE: A review of metrics - AMEE guide no. 49. *Medical Teacher* 32 (10):802–11. <https://doi.org/10.3109/0142159X.2010.507716>.
- [3] Pell G., Fuller R., Homer M., Roberts T. 2013. Advancing the objective structured clinical examination: sequential testing in theory and practice. *Medical Education* 47 (6):569–77. <https://doi.org/10.1111/medu.12136>.
- [4] Fuller, R., Homer, M.S., Pell, G. Hallam, J. 2017. Managing extremes of assessor judgement within the OSCE. *Medical Teacher*. 39(1), pp.58–66. <https://www.tandfonline.com/doi/full/10.1080/0142159X.2016.1230189>.
- [5] Homer M., Fuller R., Pell G. 2018. The benefits of sequential testing: Improved diagnostic accuracy and better outcomes for failing students. *Medical Teacher* 40 (3):275–84. <https://doi.org/10.1080/0142159X.2017.1404561>.
- [6] Homer M., Fuller R., Hallam J., Pell G. 2019. Setting defensible standards in small cohort OSCEs: Understanding better when borderline regression can 'work.' *Medical Teacher* 26 (0):1–10. <https://www.tandfonline.com/doi/full/10.1080/0142159X.2019.1681388>.

Medical Teacher and *Medical Education* are two of the leading peer-reviewed journals in the medical education field. *Medical Teacher* is the official publication of the Association for Medical Education in Europe (AMEE).

4. Details of the impact (indicative maximum 750 words)

The main beneficiaries of the underpinning research have been medical schools (nationally and internationally), medical and other trainee professionals, national regulators, and, ultimately, the public. The research has situated the Leeds team as national and international assessment experts and led to invitations to sit and advise on a range of policy panels responsible for developing or overseeing at least five key national examinations [A,B,C,D]. It has also led directly to improvements in medical assessment practices across a range of institutions in the UK and abroad [E,F,G,H]. Beyond medicine, several healthcare and allied

professions (e.g. pharmacy, nursing, dentistry, social work) continue to improve their assessment practices based on Leeds's historic and ongoing research practice and advice [I,J]. Better assessments improve the quality of practising healthcare professionals, increase patient safety, and save on resources.

(i) Impact on national regulators

Homer and Fuller have a growing set of important assessment advisory roles with UK regulators based on their proven assessment expertise and reputation underpinned by their research and its associated outputs [1-6], engaging through papers, workshops, 'meet the experts' sessions, and plenaries at international conferences to influence policy and practice.

Since 2017, Homer and Fuller have ongoing roles supporting the General Medical Council (GMC), the UK-wide statutory regulator for medical practitioners, in developing the new Medical Licensing Assessment (MLA) that all undergraduate medical students will be required to take in 2024. Their research informed the requirements for the clinical and professional skills element of the MLA, and the structure and make-up of the applied knowledge part (2019-2020), reflected in the guidance issued by the GMC to medical schools [B]. Approximately 7,000 candidates from across the world will be sitting the MLA in the UK each year, and will have to pass it to progress onto the Foundation Training Programme in the NHS. This new exam is central to assuring the public that UK medical graduates are of sufficient quality to work in the NHS.

The Chair of the GMC Medical Licensing Assessment Reference Group, which advises the GMC Council and Executive, comments that the research was '*a key influencer in determining the overall shape and balance of the programme of assessment, and the detailed approach to be used to the particular assessment methods to be used... the evidence from the Leeds group has been fundamental in shaping the approach our expert group has advised the GMC to adopt*' [A]. More specifically, the '*Leeds evidence regarding extremes in assessor judgement (e.g. Fuller et al 2017 [4]) was important in persuading us to move away from a 'centralised' model of assessment to a 'distributed' model*' (a move away from a centrally organised test to one where medical schools run their own tests but must meet a range of GMC requirements on these) '*whilst their evidence on quality assurance (e.g. Pell et al 2010 [2]) was crucial to the development of very robust quality assurance criteria to underpin such a model*' [A].

The Head of Medical School Liaison for the MLA at the GMC identifies specific contributions from the underpinning research outputs to: GMC thinking on resits [1]; the development of standards for quality assuring UK medical schools' clinical and professional skills assessments [2,6]; and evaluating the potential for using sequential testing for the applied knowledge test [3,5]. She summarizes that the '*papers have had, and will continue to have, an impact on the design, development and implementation of the MLA which, in turn, will have a direct impact on the NHS workforce. These papers provide essential evidence and background to our discussions with stakeholders about good assessment practice in undergraduate and postgraduate medical education*' [B].

In addition to work on the MLA, since 2017 Homer has also provided regular and ongoing support to the GMC to quality-assure and improve the two-part Professional and Linguistic Assessments Board (PLAB) examinations that allow international medical graduates to join the GMC register and practise medicine in the UK [C]. PLAB examines approximately 6,000 candidates each year and is essential in assuring the quality of a stream of appropriately qualified clinical staff into the NHS from overseas. PLAB1 is a multiple-choice written examination and PLAB2 is a practical OSCE. Homer is the only non-clinical member of the panel monitoring the quality and overseeing the development of this critical exam. In summer 2020, Homer carried out the analysis and wrote the [annual review report](#) on PLAB for the GMC that has informed further developments of the assessment. For example, as a result of recommendations in this report, Homer was invited to advise the PLAB1 panel on specific ways to improve their standard setting processes. He also recently began a new programme of research based on assessment data from PLAB2 that is feeding into future developments of PLAB [6]. Finally, in 2020, he also advised on the interim re-structuring of the exams in light

of COVID-19 restrictions. The GMC's Head of Assessment remarks: '[Homer, Fuller, Pell] have been influential in the design, scoring and standard setting of the [PLAB2] exam. In particular, they have advised extensively on the implementation of borderline regression [2,6] ... on variance in assessor judgements [4], defining the borderline candidate (Homer et al., 2017) and on the range of quality metrics that can be used to monitor an assessment [2] have also been influential' [C].

The Leeds research [1,3] was used by two recent reviews commissioned by the GMC (2014, 2018) to provide an evidence base for changes to their clinical assessments (PLAB and MLA respectively) [D]. Finally, since 2018 Homer has provided additional assessment expertise for the GMC on the [Curriculum Advisory Group](#), which monitors and approves all changes to post-graduate medical education curricula and associated assessment practices in the UK medical royal colleges. This work has improved these practices in the medical royal colleges via the application of modern assessment methods rooted in the Leeds research. For example, the work on setting defensible standards where there is a small cohort [6], has been used in the GMC's review of assessments in medical specialty training provided by some royal colleges and faculties, where there are only small numbers being examined [B]. Fuller led an independent [review](#) of Royal College of Radiologists examinations in 2014.

(ii) Impact on medical schools in the UK

Nationally, medical schools have used Leeds research and assessment expertise to change and improve their own assessments. Since 2013, the researchers have successfully helped leading UK medical schools, including Liverpool, Sheffield and Queen's Belfast, to improve the quality of their assessments following the practices detailed in the research [2-5]. For example, the Lead for Assessment, Kent and Medway Medical School, one of the UK's newest medical schools (f. 2020), describes the impact of the research on medical education practice as '*substantial and profound*' and explains: '*The newly established Kent and Medway medical school in Canterbury, Kent has relied heavily on the published research... [2,3,5,6] to shape our assessment and remediation policy for this new medical school*' [E]. The Dean of Brighton and Sussex Medical School states: '*[the] School has undertaken a root and branch review of its assessment strategy... supported by Dr Richard Fuller... this review has drawn on the published literature in this area [1,3,5] to guide our decisions*' [F].

To date, at least eight established and new medical schools (out of 33 in the UK) have moved or are moving to a sequential model of assessment following Leeds's example and resulting from the underpinning research [3,5] and related knowledge exchange activities. These include Liverpool, Hull York, Sheffield, Aberdeen, Brighton and Sussex, Kings College London, Leicester and new medical schools coming on stream (e.g. Kent and Medway) [E,F,G]. The research estimates that adopting this assessment model saves a large medical school around GBP90,000 per year [3]. The Lead for Assessment at Kent and Medway Medical School continues: '*We have incorporated the use of sequential testing in OSCEs and extended its use to our knowledge testing. We also share the annual remediation (longer-term) view of deep remediation for students who are unable to pass their sequential assessments*' [E]. On the rationale for moving to a sequential assessment design, the Dean of Brighton and Sussex Medical School highlights changes based on three findings from the underpinning research: '*One of the changes to the BSMS assessment strategy is to introduce sequential testing to our OSCE examinations...to reduce unnecessary testing burden on students [5], ... to support students who do not reach the passing standard by providing them with a substantial period of remediation rather than the few weeks presently available [1] ... [and] to improve the reliability of the OSCE examination [3]*' [F].

(iii) Impact on medical and other healthcare schools internationally

The research-informed assessment expertise has resulted in the team advising on assessment quality and design issues across the world. This includes face-to-face visits and/or electronic communication with a wide range of leading educational organisations, including in Europe (Karolinska Institut, Minho, Portugal); South East Asia (International Medical University in Malaysia and Nanyang Technological University in Singapore); the Middle East (various institutions in Kuwait and UAE); North America (University of Arkansas for Medical Sciences, Faculty of Veterinary Medicine, University of Calgary); and Australia

(supporting development of OSCEs in Australia for internationally qualified nurses). The Pro-Vice Chancellor for Education and Institutional Development at International Medical University, Kuala Lumpur (Malaysia's leading private medical and healthcare university) remarks that *'expertise from Leeds was valuable to provide contextual solutions to assessments challenges at IMU. [Fuller and Pell] ran workshops and conducted reviews at IMU helping to improve assessment of clinical skills and post exam analysis at IMU'*. Specifically, outputs from the underpinning research [4-6], were used to benchmark best practices in assessments [H].

Beyond the training of doctors, regarding a recent large-scale national pilot of a nursing OSCE that she led, the Dean of Research, College of Science, Health, Engineering and Education, Murdoch University, Australia (formerly based at Monash University where the work was co-ordinated) explains that: *'Based on their international reputation and expertise in OSCEs, we invited ... Homer and ... Fuller to be part of an international advisory panel at Monash University to help us develop, pilot and evaluate OSCEs for internationally qualified nurses for the Nursing & Midwifery Board of Australia... They were instrumental in advising us about the interpretation of evaluation data with respect to a range of quality metrics for the OSCE ... leading to us developing a reliable, cost-effective and acceptable OSCE' [I]*. This successful pilot was the basis for the new assessment model for overseas qualified nurses and midwives adopted by the Nursing and Midwifery Board of Australia in March 2020. In Canada, the Co-Chair of Faculty of Veterinary Medicine at the University of Calgary reports being 'inspired to research the topic [sequential testing]' and subsequently 'instituting a similar type of OSCE format' after encountering the research and corresponding with Homer and Pell [J].

The research has enhanced assessment practices, design of assessment instruments and psychometric quality-assurance approaches in medical and allied healthcare educational institutions and regulators, nationally and internationally. This work safeguards the quality of a range of healthcare practitioners entering and progressing in the workforce, saves resources, and enhances patient safety in the long term.

5. Sources to corroborate the impact (indicative maximum of 10 references)

[A] Chair of Medical Licensing Assessment Expert Reference Group, GMC, in relation to 4(i). Letter of corroboration dated 21 November 2019.

[B] Head of Medical School Liaison for the Medical Licensing Assessment, GMC, in relation to 4(i). Letter of corroboration dated 30 December 2019; GMC (2019), Requirements for the Medical Licensing Assessment Clinical and Professional Skills Assessment (para. 4,21).

[C] Head of Assessment, GMC, in relation to 4(i). Letter of corroboration dated 12 August 2020.

[D] GMC-commissioned reviews, in relation to 4(ii). McLachlan, J. et al (2014), [Developing an evidence base for the Professional and Linguistic Assessments Board test](#) (pp. 29,30,57); Burford, B. et al (2018), [Best Practice in the Assessment of Competence: A Literature Review](#) (pp. 49,50,55).

[E] Lead for Assessment, Kent and Medway Medical School, Canterbury Christ Church University and University of Kent, in relation to 4(ii). Letter of corroboration dated 29 November 2019.

[F] Dean of Brighton and Sussex Medical School, University of Brighton, UK (and colleagues), in relation to 4(ii). Letter of corroboration dated 11 September 2020.

[G] Examples of online documents evidencing adoption of sequential OSCEs at UK medical schools, in relation to 4(ii): [Liverpool](#); [Hull-York](#); [Brighton and Sussex](#); [Aberdeen](#).

[H] Pro-Vice Chancellor for Learning and Teaching, International Medical University, Kuala Lumpur, Malaysia in relation to 4(iii). Email dated 10 August 2020.

[I] Dean of Research, College of Science, Health, Engineering and Education, Murdoch University, Australia, in relation to 4(iii). Letter of corroboration dated 8 September 2020.

[J] Co-Chair of Faculty of Veterinary Medicine, University of Calgary, Alberta, Canada, in relation to 4(iii). Letter of corroboration dated 8 September 2020.