The Applied Knowledge Test – theory and practice

Dr Marco GRECH

ABSTRACT

The Applied Knowledge Test (AKT) forms part of the summative assessment for the Membership of the Malta College of Family Doctors (MMCFD). Candidates who are successful in the summative assessment and who have successfully finished the Specialist Training Programme in Family Medicine are awarded the MMCFD and the MRCGP[Int] on the basis of a tripartite agreement in place between the Government of Malta, the Malta College of Family Doctors and the Royal College of General Practitioners. This article looks at the local setup of the AKT. It explains the whole process from item writing, to piloting, blueprinting and standard setting. The article also attempts to explore the theory behind the AKT that underpins it as a reliable, valid, educational, cost-effective and acceptable mode of assessment within Miller's pyramid of clinical competence.

Keywords

Applied knowledge test, assessment

INTRODUCTION

The Applied Knowledge Test (AKT) forms part of the summative assessment for the Membership of the Malta College of Family Doctors (MMCFD). The overall purpose of this final summative assessment is to assess the competence of general practice (GP) trainees who have finished or are in the last six months of the Specialist Training Programme in Family Medicine (STPFM). Having achieved this level of competence, candidates are awarded the Membership of the Malta College of Family Doctors. This, together with the certification of completion of training, enables the candidates to apply to the Specialist Accreditation Committee for listing as Specialists in Family Medicine. It also enables candidates to be awarded with Membership of the Royal College of General Practitioners (MRCGP [Int]) according to a tripartite agreement currently in place between the Government of Malta, the Royal College of General Practitioners (RCGP) and the MCFD.

THE APPLIED KNOWLEDGE TEST

The AKT is a 3-hour 200 multi-choice question examination aimed at testing the application of knowledge in the context of Maltese Family Medicine. There are no true-or-false questions and therefore negative marking is not applied. The AKT attempts to assess both clinical and non-clinical aspects of family medicine, with assessment of medicine related to general practice such as general medicine & surgery, medical specialties (e.g. dermatology, psychiatry, geriatrics), surgical specialties (e.g. ENT, ophthalmology), women's health and paediatrics. Critical appraisal and research methodology related questions are also included. Each question is intended to explore a topic about which an ordinary general practitioner (GP) in Malta is expected to have a working knowledge.

The questions in the AKT are designed to assess knowledge about evidence-based current best practice rather than local practices. Questions are written by a group of practising local GPs who are offered training in AKT writing by the MCFD. These writers bind themselves by a confidentiality agreement. All test items in the AKT are based on the MCFD Curriculum blueprint. All questions have to be referenced. This facilitates the verification of answers and the updating of the questions in the future. After an initial feedback by the AKT lead, all questions are peer reviewed within the AKT writers' group and refined as necessary. Renowned reference sites are used when writing questions. These include the National Institute for Health and Care Excellence (NICE) and Scottish Intercollegiate Guidelines Network (SIGN) guidelines, the British Medical Journal (BMJ), the British Journal of General Practice (BJGP), Medline, and the British National Formulary (BNF). Use is also made of a number of online resources such as the RCGP Essential Knowledge Updates, BMJ Learning, and the Clinical Knowledge Summaries (now clarity.com). Following this process, questions are stored in a bank ready for selection and inclusion in an exam paper.

Questions in the AKT take one of two forms: the Single Best Answer (SBA) or the Extended Matching Question (EMQ). In SBA questions, a stem presents a clinical scenario or a factual statement. This is then followed by a list of five possible options. Only one option can be chosen and the candidate will have to decide on the "most appropriate answer". (Elfes, 2011)

An Extended Matching Question is a selected response item in which the item stem has been extended, usually, to a short clinical vignette or scenario and the choices have been extended to include all potentially acceptable ones for the clinical problem or issue that is being addressed by the item (Jolly, 2014). Pictures may form part of either of the two types of question.

All GP Trainees who

- have successfully completed the three-year Specialist Training Programme in Family Medicine (STPFM),
- will be completing the three-year STPFM programme within 6 months from the date of the examination, or
- have failed previous sittings of the AKT component as stipulated by the regulations

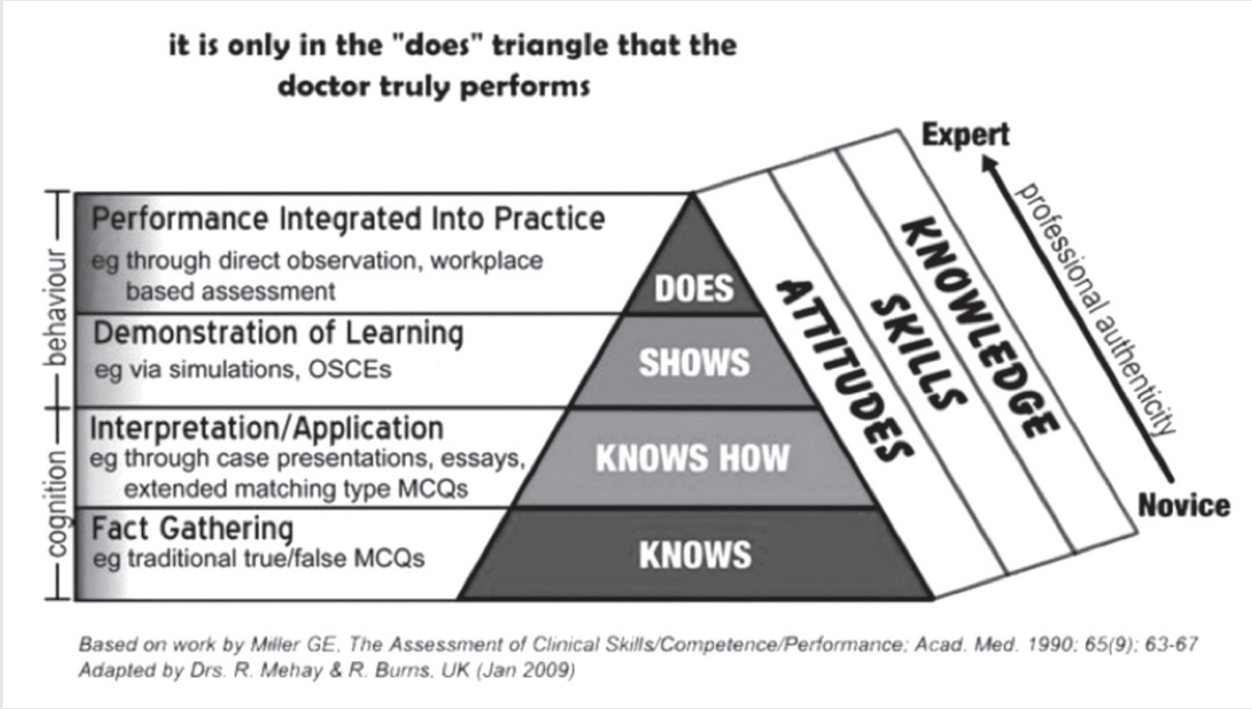
are eligible to sit for the AKT and Clinical Skills Assessment (CSA) components of the MMCFD examination. (Malta College of Family Doctors, 2013)

The examination is usually held at the Malta Medical School. The whole process is monitored by the MCFD's Quality Assurance officials. This ensures transparency and that the correct procedure (e.g. that the paper is sealed before being opened) is being followed throughout.

Standard setting involves the definition of a clear standard below which a trainee GP would not be deemed fit to practice independently (Wass et al., 2001). Such a standard is set locally using the Angoff method wherein a group of 9 practising GPs, comprising a healthy mix of experienced and newly qualified GPs, come up with the cut-off point after analyzing every question in the AKT paper in detail. These GPs are reminded in every session that the established cut-off point would identify the "minimally competent GP". Essentially this group is asked to individually rate the probability of a borderline candidate passing an individual question in the test. Any wide variations are resolved after discussion within the group. This is a very laborious process which takes a number of sessions but is essential in producing a fair outcome for all parties. The Angoff group sessions are held before the sealed papers from the AKT exam are corrected, thereby eliminating the possibility of the introduction of bias in the standard setting procedure.

The correction of the paper is done by hand using answer sheet templates after the Angoff procedure has

Figure 1: Miller's prism of clinical competence (aka Miller's Pyramid)



been finalised. Each paper is corrected by two separate examiners and any discrepancies in the marks awarded by the two examiners are reviewed by a third examiner. Both the standard-setting Angoff group sessions and the correction of the papers are closely monitored by the College's Quality Assurance officials.

The pass mark is then set using the cut-off score that is the product of the Angoff process and the Standard Error of Measurement that is a statistical function of the set of scores obtained by the candidates in the AKT examination.

THE THEORY OF ASSESSMENT AND THE AKT

Assessment drives learning (Wass et al., 2001). Formative assessment is used to promote learning. The feedback received by trainees during their training should be aimed to build their knowledge and skills. Assessment needs also to have a summative function. It is only thus that a doctor can be certified as being fit to practise, thereby satisfying the demand by the profession and the public for assurance that doctors are competent.

The AKT aims to assess the application of knowledge, not just the recall of knowledge, in a wide variety of scenarios. This would correspond to the "knows how" level in the Miller's Prism of Clinical Competence (see Figure 1) (Wass et al., 2001). The other components of assessment leading to the MMCFD cover other levels of this pyramid. The Clinical Skills Assessment covers the "shows how" level, whereas the Workplace-Based Assessment covers the "does" level of competency.

In his seminal work, van der Vleuten (Van der Vleuten, 1996) looks at the characteristics of a good assessment system. Van der Vleuten suggested that reliability, validity, educational impact, cost effectiveness and acceptability are to be considered in the construction of an assessment system.

Reliability

Reliability refers to the reproducibility or the consistency of a test. (Wass et al, 2001). It indicates the ability of a test to be replicated under the same conditions. Reliability can be seen as the ratio between subject variance (what we are trying to measure in an exam) and the subject + error variance. The reliability coefficient measures what percentage of the variance is due to true differences between candidates and what percentage is due to error (General Medical Council, 2010). It can therefore be improved by increasing the variance between candidates relative to error variance. Cronbach alpha is the most widely used reliability measure. The coefficient gives a value between 0 and 1; the latter value would reflect the perfect test. A cut-off of 0.8 is traditionally taken as a benchmark of reliability. All assessments have an inherent element of error which can never be removed completely, though much can be done to reduce this level of error to the minimum possible e.g. by eliminating ambiguous questions and by intensive examiner training (Tighe et al., 2010).

One can also calculate the effect of any error that remains. The Standard Error of Measurement (SEM) provides the confidence interval around the pass mark. The smaller the SEM, the more accurate is the assessment that is being made. Some have suggested that the SEM is a more appropriate measure of quality for postgraduate medical assessments than reliability (Tighe et al., 2010). This is because the reliability coefficient can be artificially inflated by having a greater number of very weak or very strong candidates sit for the exam. This will increase the standard deviation and as a result the reliability will apparently be higher. When examinations have a very small number of candidates the risk that reliability is distorted by an unusually high, or low, spread of candidate ability is greater. The SEM's main use is in the proper identification of borderline candidates – those whom the examination has not been able to confidently place on one side or the other of the pass mark (Postgraduate Medical Education and Training Board, 2007 cited in Tighe et al., 2010). A low SEM would indicate a higher accuracy achieved in the classification of the cut-off point.

Validity

Validity is defined as the extent to which the competence that the assessment claims to measure is actually being measured (Schuwirth and van der Vleuten, 2006). Two main types of validity are considered: content validity and construct validity.

The *content validity* in the AKT relates to whether the assessment covers the whole spectrum of what has to be tested, which in the local scenario is the Curriculum of the MCFD. It is the role of the Assessment Team to ensure that the AKT paper covers the whole blueprint of the curriculum. As assessment drives learning (Eraut, 2004 and van der Vleuten and Schuwirth, 2005) this wide representation of the blueprint conveys an educational message to the trainees of what is needed to master the test.

A *construct* is defined as a personalised psychological characteristic that cannot be observed directly but which

is assumed to exist (Schuwirth and van der Vleuten, 2006). So in construct validity (also known as indirect validity) we are trying to assess whether the assessment scores align with our expectations about the type of competence we are trying to assess. Therefore, in a medical problem-solving test with a good construct validity one would expect that people who solve problems more expertly to outperform those who are less good problem-solvers (Schuwirth and van der Vleuten, 2006).

Other types of validity exist and are sometimes referred to. Perhaps in the future more impressive evidence for the AKT will emerge from studies, which to date are not available, about the extent to which the AKT predicts later performance. (Metcalfe, 2012)

Educational impact

Evidence shows that assessment has a major impact on students' study behaviour (Jolly, 2014). The content, format, scheduling and regulatory structure of assessments can have a positive or negative effect on the intrinsic and extrinsic motivation for learning of trainees (Schuwirth and van der Vleuten, 2006). Some summarise this as "students don't do what you expect, students do what you inspect". Therefore assessment can be used to influence the students' learning in several ways. Having the questions tied to the curriculum blueprint helps ensure that candidates read about a variety of subjects during their studies. Studies may be needed to assess the candidates' reading behaviour when preparing for the AKTs and how this compares to the reading behaviour adopted when preparing for the CSAs, for example.

To be eligible to sit for the AKT in Malta, the GP trainees would have to have finished, or are in the last six months of, the Specialist Training Programme in Family Medicine. One session per calendar year is held locally. This contrasts with the possibility in the UK of GP trainees sitting for the exam in one of three sittings throughout the last two years of training, thereby having the facility to choose the ideal time to sit for the examination (Metcalfe, 2012). It is evident that the MCFD lacks the resources to organise this any time soon. One hopes that the capacity-building exercise being encouraged by the current MCFD Council bears fruit in this respect as well.

Cost effectiveness

The cost effectiveness of an assessment is a compromise between the information gained and the resources required (van der Vleuten, 1996). The cost for the candidate to sit for the MCFD AKT exam in 2014 was set at €500. Costs incurred in running the exam include remuneration of writers, examiners, invigilators, members of the Angoff group and members of faculty, together with printing, secretarial services and other minor sundry expenses.

A difficulty arises in assessing the cost-effectiveness of the AKT exam in isolation. One would rather look at it as part of the whole MCFD exam considering that some of the costs are shared. However it is generally accepted that an MCQ examination is considered as one of the most cost-effective and reliable examinations to assess the "know" and "knows how" levels on the Miller's pyramid (Metcalfe, 2012).

Locally, the examination delivery and correction is still paper-based. Other centres administering similar examinations have switched to computer-based technology (Metcalfe, 2012). The introduction of such technology could introduce a number of advantages such as:

- a reduction in human resources needed, e.g. examiners, invigilators;
- improved efficacy in the marking and analysis of the examination;
- a reduction in the human error possibility, e.g. while correcting;
- feedback for individual candidates and for the whole cohort become easier and quicker.

On the other hand the introduction of such technology might create some disadvantages such as:

- the introduction of bias between candidates on the basis of their technological abilities;
- higher design costs;
- costs of hardware and networks and the maintenance thereof (Metcalfe, 2012);
- the reduction in cost-effectiveness caused by the limited number of local candidates.

Acceptability

Van der Vleuten proposes that the beliefs, opinions, and attitudes of both examiners and examinees must be considered in choosing and designing assessments in order to ensure that there is no threat to the survival of the assessment (Postgraduate Medical Education and Training Board, 2008).

No studies have been conducted locally to assess the acceptability of AKTs to examiners. However it is well known that the AKT process is lengthy, requiring

time to research questions which will then need modification, peer-reviewing, re-modification after reviewing, categorisation before inclusion in the bank and standard setting. Questions also need to be continually updated with the latest guidelines. Item analysis after the exam is also another time-consuming exercise in which all items in the exam are analysed for discrimination and improved as necessary.

On the other hand, evaluation among candidates indicates a general widespread acceptability of the AKT exam. After the 3-hour examination, the candidates dedicate quite some time to fill in the evaluation form. This shows their high degree of interest and appreciation of the exam process as a whole.

The organisational and logistical aspects of the examination process were all highly rated. A marked improvement has also been noted lately regarding the candidates' satisfaction with the quality of the picture booklet – all candidates scored Likert 4 or 5. There was a mixed (but mostly positive) response about the spread of AKT questions as reflecting the breadth and reality of family practice in Malta. Despite all candidates finishing on time, a small minority of candidates felt that not enough time was allocated or considered the paper unfair. (Malta College of Family Doctors – AKT Exam 2014)

CONCLUSION

The strength of the MCFD assessment programme stems from combination of the formative assessment in the Work-Place Based Assessment (which promotes continuous learning through continuous feedback) and the use of different summative assessment methods each assessing different competencies in the commonly described educational theory model of Miller's pyramid. This triangulation helps increase the usefulness of AKTs in assessment as part of a complete picture of the performance of the trainees. (van der Vleuten and Schuwirth, 2005)

Reference

Elfes, C., 2011. Introduction to the Applied Knowledge Test. *InnovAiT: The RCGP Journal for Associates in Training*, Vol4, No.11 pp.667-668.

Eraut, M., 2004. A wider perspective on assessment. *Medical Education* Vol. 38 pp. 800-804

General Medical Council, 2010. Reliability Issues in the assessment of small cohorts [online] Available at: http://www.gmc-uk.org/Reliability_issues_in_the_assessment_of_small_cohorts_0410.pdf_48904895.pdf . [Accessed 3 November 2014]

Jolly, B., 2014. Written Assessment. In: T. Swanwick, ed. 2014. *Understanding Medical Education Theory Evidence, Theory and Practice*. Second Edition. Chichester: John Wiley and Sons Ltd, Chapter 19.

Malta College of Family Doctors, 2013. *Information and Regulations for the Membership Examination of the Malta College of Family Doctors (MMCFD)* [online] Available at: http://mcfd.org.mt/wp-content/uploads/Information-and-Regulations-for-the-Membership-Exam.pdf [Accessed 5 October 2014]

Malta College of Family Doctors – AKT Exam 2014. *Candidates' Evaluation Sheet* – not published.

Metcalfe, N.H., 2012. Testing the Test: an analysis of the MRCGP Applied Knowledge Test as an assessment tool. *Education for Primary Care* Vol.23, pp. 13-18

Postgraduate Medical Education and Training Board, 2007. *Developing and Maintaining an assessment system – a PMETB guide to good practice*. London. As cited in Tighe, J., McManus, I.C., Dewhurst N.G., Chis L., and Mucklow J., 2010. The Standard Error of Measurement is a more appropriate measure of quality for postgraduate medical assessments than is reliability: an analysis of MRCP(UK) examinations. BMC Medical EducationVol. 10 pp. 40-48.

Postgraduate Medical Education and Training Board, 2008. *Standards for Curricula and Assessment*. PMETB: London

Schuwirth, L.W.T., and van der Vleuten, C.P.M., 2006. *How to design a Useful Test: the principles of assessment*. ASME: Edinburgh

Tighe, J., McManus, I.C., Dewhurst N.G., Chis L., and Mucklow J., 2010. The Standard Error of Measurement is a more appropriate measure of quality for postgraduate medical assessments than is reliability: an analysis of MRCP(UK) examinations. *BMC Medical Education* Vol. 10 pp. 40-48.

Van der Vleuten, C., 1996. The assessment of professional competence: developments, research and practical implications. *Advances in Health Science Education* Vol 1. Pp. 41-67.

Van der Vleuten, C.P.M., and Schuwirth L.W., 2005. Assessing professional competence: from methods to programmes. *Medical Education* Vol. 39 pp. 309-317.

Wass, V., Van der Vleuten, C., Shatzer, J. and Jones R. 2001. Assessment of Clinical Competence. *The Lancet* Vol. 357 pp. 945-949.

Dr Marco GRECH

MD, MSc (Ulster), Cert. Diab. (ICGP), MMCFD

Assessment Lead, Malta College of Family Doctors

Email: marcogrech@yahoo.co.uk

Acknowledgements

Appreciation goes to Dr Doreen Cassar and Dr Patricia De Gabriele for their pioneering work in assessment and their encouragement and support to the author in his role.