Ottawa 2010 Consensus Statement
Technology-Enabled Assessment of Health Profession Education
Working Draft

Theme Group Members
Zubair Amin MD MHPE (lead); National University of Singapore, Singapore
Jack Boulet PhD; Education Commission for Foreign Medical Graduates, Philadelphia, USA
David A. Cook MD MHPE; Mayo Clinic College of Medicine, Rochester, Minnesota, USA
Rachel Ellaway PhD; Northern Ontario School of Medicine, Ontario, Canada
Ahmad Fahal MBBS, MD, MS, FRCS; University of Khartoum, Sudan
Roger Kneebone PhD, FRCS and FRCGP; Imperial College, London, UK
Moira Maley PhD; Rural Clinical School; The University of Western Australia, Australia
Doris Ostergaard MD; Danish Institute for Medical Simulation, Copenhagen University, Denmark
Gominda Ponamperuma MBBS, MMEd, PhD; University of Colombo, Sri Lanka
Andy Wearn MBChB, MMedSc, MRCGP; University of Auckland, New Zealand
Amitai Ziv MD; Israel Centre for Medical Simulation, Israel

1. Introduction

We live in a world of technology. It has become difficult to remember a world without high quality, synthesized electronic information at our fingertips. But the rise of technology was not smooth; it didn't deliver on its promise immediately; systems were slow, information was not summarized and the trustworthiness of data was uncertain (Hersh 1999), and many would argue that this is still the case. Yet, the last decade has seen technologies for learning, clinical practice, and assessment become an integrated part of our professional lives. A fundamental challenge for us is how we make appropriate and well-informed choices about the ways in which we use technologies in support of assessment.

This Consensus Statement represents the perspectives of this group along with inputs from other Theme Groups and conference participants. The recommendations outlined here are intended to be pragmatic and useful to a wide global audience: examiners, administrators, researchers, and policy makers involved in assessment across diverse health professions. Our focuses are on information and communication technologies (ICTs), simulation and simulators (mechanical, electronic, and human) including electronic medical records (EMRs), digital archiving and management systems, handheld devices, and assessment management tools used to improve assessment of health professionals from students to practitioners.

In an environment of multiple stakeholders and agendas, differing views on the purpose and priorities of assessment are expected. (Epstein 2007, Amin, Chong, Khoo 2006; Norcini et al 2010) An assessment system that works for administrators may ensure an efficient workflow but may neglect issues which are important to learners (e.g., support for learning, fairness) or their teachers (e.g., validity and reliability). In situations with conflicting priorities, we strive for a balanced view that takes into account the needs of all stakeholders and recognizes that assessment is closely linked to learning and clinical practice.

Annex A lists the definition used in this paper.

2. Technology in Context

A central characteristic of technology is change; every new form tends to subvert both its predecessors and itself in suggesting future forms (Graham, 1999). Nevertheless, not everything changes when technology is introduced into a new situation. We can therefore consider three essential aspects of change associated with use of technology in assessment – transmediation, innovation and prosthesis:

- *Transmediation:* refers to the process of taking existing information, practices, and tools and moving them to a new medium. For example, the traditional end-of-course paper-based examinations involving whole classes sitting down together are being transmediated to an on-screen medium, with much of the marking automated and the results posted within the course learning management system instead of on a public noticeboard. Technology in this context principally involves taking existing tools and practice (e.g., paper-and-pencil based MCQ) and moving them to a new medium (online web applications and databases).
- *Innovations:* are forms and processes that could not and did not exist without technology. Recent innovations in assessment include various uses of simulation (Margolis et al, 2004; Gesundheit et al, 2009) as well as dynamic media such as interactive images and video. The relative extent of transmediation and innovation indicates how much of the original has been translated and how much is truly original. The extent of real innovation is typically small as we tend to make small rather than giant leaps. However, by incorporating established practices we may inadvertently introduce artifacts and enforce inappropriate orthodoxies in the design and use of the technology in question (Scarborough and Corbett 1992). For instance

online exams are often still called 'papers' and unnecessarily follow the limitations of pulp-based examination materials.

- *Prosthesis:* is about extending human action beyond the limits of the physical self to allow us to do things faster, more accurately and in more places simultaneously and to work with things too big or small, or too fast or slow to handle unassisted. For instance, technology can significantly extend and enhance the workflows and logistics associated with assessment including creating assessment materials, managing security and marking (at least where human intervention is not required). Indeed, in many cases efficiency and quality control is the primary reason for institutions to adopt e-assessment (Bennet 2005).

3. Current Status

3.1 Technology in Health Professional Education

Both medicine and education are intrinsically technology-enabled phenomena (Reiser 2009; Economist 2010). Medicine is dependent on systematized knowledge, as well as on its tools, drugs and other paraphernalia. Becoming a doctor is in many ways synonymous with becoming an ethical technocrat with humanistic goals of care, enlightenment and supporting civil society; someone whose authority significantly comes from their appropriate use and control of technology (e.g., tools, medicines, devices, and knowledge repositories). Similarly, education is founded on the systemization of learning, again significantly structured around technologies such as books, classrooms, devices and simulation (Kress 2010).
Information technology is clearly fundamental to any healthcare system and is currently manifested explicitly in the form of electronic medical records, electronic ordering systems, Picture Archival and Communication Systems (PACS) and billing systems, and more generally through PubMed, online journals and databases and handheld and mobile technologies. Healthcare professionals now regularly use IT as a core part of the care-giving processes as well as in their research, evidence-based practice and professional development. The uptake of new technologies remains rapid with two-thirds of physicians and 42% of the public using smart phones as of late 2009 (CHCF 2010). The creation of applications related to health and health care is also moving quickly; as of February 2010, there were nearly 6,000 medical applications available from Apple's iTunes App Store (CHCF 2010).

Higher education is also suffused with technologies including PowerPoint presentations and learning management systems. Whether we approve or not, students' research is as much about using Google and Wikipedia as it is about using an institutional library. While much of the ICTs used in HPE are fairly generic, there are a number of healthcare-specific technologies, in particular applications of simulation ranging from low-fidelity task trainers to high-fidelity mannequins (Issenberg, 2005).

Clearly, the applications and significance of technologies in healthcare and education are broad-based, inherent, and pervasive (Greenhalgh, 2001; Ellaway and Masters, 2008). It is in this context that technology in assessment of medical education must be seen, in part because it defines what is normative, acceptable and sustainable and in part because any good assessment in HPE should faithfully reflect the context and environment of learning clinical care.

3.2 Technology and the Continuum of Learning, Clinical Practice and Assessment

Clinical care, and the technologies that are used within it, gives rise to three environments of interest – the clinical practice or real life environment, the learning/training environment, and the assessment environment. One could argue that the more we can ensure a continuum between the three environments in terms of technology use, the more we can assist learning and ensure a higher level of context to the assessment process and to its validity (Ellaway, Kneebone et al 2009).

The link between the three environments can be illustrated by the following examples:

- In carotid artery stenting, a specific patient pathology is evaluated by CT-Angiogram. CT digital data of that patient is transmediated into a carotid artery stenting simulator, enabling the practitioner to practice in the context of a specific patient, with its specific risks and procedural challenges. Only after the practitioner demonstrates a satisfactory proficiency and safety level, the procedure is conducted on the real-patient in the real-life. This establishes the continuum between clinical context into training, assessment, and back to the real-life environment (Black 2006, Black 2010).
- Electronic medical records (EMRs) can be used to train and assess health professionals within the local clinical context (Keenan et al, 2006). A simulated environment can be customized and personalized to include EMRs, enabling trainees to perform a clinical task on simulators/simulated patients or on a combination of both (the hybrid model). The task is carried out within the context and driven by real-life cases, patients, and clinical challenges. Technology here maintains the continuum between a) a real-life environment; b) a fully simulated environment (utilizing simulated patients/simulators); and c) a combination of both– the "simulation-in-situ" concept where simulators or simulated patients (or both) are employed within a real-life environment.

3.3 Applications of Technology in HPE Assessment

Although the use of technology in assessment is not new (Bradley 2005;Tekian 1999) and various forms of technology have been used in assessment for many years, major developments of technology in HPE assessment are largely centered on computer-based assessment, use of simulation and simulators (Norcini and McKinley 2007), and management of assessment processes. Not surprisingly, some of the strongest evidence supporting the use of technology comes from these areas (Norcini et al 2010).
Computer-based Assessments: Over the past two decades, improvements in computer technology have led to many enhancements in item and test construction, test delivery and scoring. The use of paper and pencil multiple choice examinations has gradually been transmediated into computer-based delivery of test content, often over the internet via secure, encrypted, connections. Provided the item pool is sufficiently large and a detailed blueprint exists, automated test construction software can be used to generate multiple test forms (Norcini McKinley 2007). Moreover, the use of computers enables rapid scoring, including the generation of tailored feedback and adaptive testing. The use of technology also allows for the construction of computerized cases where those being assessed are tasked with managing a patient (or patients) in simulated real time on the computer (Dillon, Boulet, Hawkins, Swanson 2004).

Simulation and Simulators: There have also been many innovations in simulator technology, including part task trainers and various electromechanical mannequins (Norcini, McKinley 2007). As these technologies expand, so does the available assessment domain. Unlike traditional standardized patient assessments, where general interviewing skills are typically assessed, simulators allow for the modeling of a multitude of abnormal physical findings. Part-task trainers can also be combined with standardized patients to allow hybrid model of assessment of procedural (e.g., suturing using a skin pad) and communication skills (Kneebone 2002). Crisis events can also be modeled, allowing healthcare teams to be evaluated in realistic environments (Wong 2002, Sica 1999). Distributed simulation using portable, low-cost, and highly immersive environment offers a new avenue of testing clinical skills in authentic setting (Kneebone 2010). Overall, advancements in simulator technology have opened the door for more authentic assessments that can be used to evaluate a much wider range of skills (Issenberg et al 2005)

Management of Assessment Processes: Although this is a major (maybe 'the' major) application of technology in assessment it is often overlooked; perhaps because students are not directly involved, and it is in support of the logistics, administration and analytical aspects of assessment. It is often in this area that the greatest benefits are to be found in terms of improved efficiency, tracking and quality assurance of assessment (Ellaway and Masters, 2008). Examples of technology in assessment resource management include item banking, indication of plagiarism, data monitoring and reporting, analysis of results, remote tracking and telemetry.

Annex B summarizes some of the contemporary usages of technology in assessment.

3.4 Benefits of Using Technology in Assessment

The benefits arising from the application of information technologies in the assessment of HPE are well described (Issenberg et al 1999, Gordon JA 2001, McGaghie 2006, Ziv 2003, Kneebone 2009). Particular benefits of using technology in assessment include:

- *Learner-centeredness:* Simulation technology creates a learner-safe and learner-centric environment where the mistakes are 'forgiven' and feedback is immediate (Issenberg et al, 2005). It is now possible to assess a surgeon's dexterity from the telemetry generated by their manipulation of a probe or scalpel (Fried 2004); a trainee's ability to stabilize a patient in cardiac arrest (using a high fidelity mannequin) can be assessed from the adherence (or variance) from a defined scenario protocol (Boulet, Murray 2003); or a candidate's decision-making abilities from their success in 'playing' a game-like virtual patient case. In all such cases the primary focus is the learner, not the patient as in real-life, making the assessment learner-centric process.
- *Assessment of Broader Domains of Competencies:* Technology allows us to assess domains of physicians' competencies that could not be tested before. For example, we can assess how a group of people performs in a given situation rather than individual performance (Dev, Youngblood et al, 2007). Simulation can provide an experience that gets close to real practice and can be used for high stakes (Boulet 2008) and rare but important clinical events that are critical for patient-safe care (Garden 2002). In clinical settings, technology can compress the time frame for decision making or add a second tier of validation on the data being generated, therefore broadening the educator's brief to implement professional skills training that accommodates not only clinical reasoning from direct patient examination, but discerns genuine data from artifacts (Schuwirth 1996).
- *Assessment of Performance:* Technology can support the move from "shows how" to the "does" level in Miller's pyramid (Miller 1990). For instance, video recordings of real patient encounters can be used in performance assessment supplemented by an oral assessment or a written reflective commentary. Covert, but consented, recordings or 'mystery shopper' events (actors posing as patients as part of routine practice) can be used to help assess performance at the "does" level. Furthermore, auditing clinical databases and patient outcomes can be used to monitor progress and maintenance of competence among practitioners (Scalese, Issenberg 2005; McGaghie 2007). Some of the better-established Quality Assurance (QA) processes in health care can also be used in performance assessment. For example, in some developed world settings, data mining of EMR is now routinely employed to determine clinicians' performance in terms of outcome measures like morbidity and mortality, or process measures such as "adherence to clinical guidelines" (Shavit 2007; Farfel 2010)F. The evolving technology enables these assessment approaches to be applied in earlier stages of HPE training as well as in formative and summative assessment.
- *Efficiency and Simplicity of Assessment Processes:* Improved assessment processes may potentially benefit all stakeholders involved in assessment. For example, from a candidate's perspectives, a shorter turn-around time between the examination and release of the result is helpful for a number reasons: it alleviates anxiety; it allows for more direct cognitive inks between the assessment event and the results of personal

performance; it allows more time for remediation if necessary; and it extends the actual learning time. Assessment authoring tools allow faculty authors to create, edit and repurpose assessment items (typically but not exclusively in the form of questions). Logistics and examination delivery tools allow faculty assessors to set rubrics, analyze psychometrics, validate examination, check plagiarism, and support quality assurance and audit processes. A tracking system with centralized data reporting system on candidates' clinical experience gives important information about the quality of clinical training to the curriculum administrators and allows constant monitoring of progress. At a more fundamental level, these kinds of tools help ensure that there is good alignment between actual learning and the curriculum and they assist in assessment blueprinting and quantitative evaluation of assessment across one or more programs.

Despite these potential benefits, the use of technology in assessment has not been as widespread as its application in support of learning (Bennet 2005; Tucker 2009). It is the collective opinion of the Group that judicious use of technology can greatly improve the assessment practices across the spectrum of HPE by improving learner-centeredness, by including important elements of professional performance in assessment, and by streamlining and improving the management of assessment processes in general.

4. Challenges

The arguments favoring use of technology in learning and assessment in HPE are well described (Issenberg et al 1999; Ziv 2003). However, meaningful application and adoption of technology in assessment also requires an understanding of the challenges and threats to the validity of assessment arising from technical affordance, the breakneck pace of technical development (and redundancy), and inequities of resource and access.

4.1 Validity Threats

The ultimate purpose of assessment is to make some kind of inference or decision about the candidates based on their performance. Defensible conclusions (strong inferences) require valid psychometrics or at least well-grounded non-psychometrical models of performance. An informed assessor's responsibility is to be aware of potential threats to the validity that may results from the use of technology-mediated assessment and ways to rectify them.

4.1.1 Construct under Assessment

Whenever we conduct assessment we attempt to make inferences about the learner – inferences about their knowledge, attitudes, general competence, communication skills, and so forth. This intended inference is called a construct. If the construct is not well structured within the design then using a technology for assessment may, for instance, measure the candidate's ability to use the technology rather than evaluate the actual clinical performance; i.e. the intended construct. Obviously, if the purpose of the assessment is to assess candidate's ability to the use of technology (such as working with an EMR) then the construct is in part the use of technology (Shachak 2009). However, we tend to use technologies, such as a human patient simulator, to approximate expected performance in another setting (a live patient). In these situations, the candidate's comfort and skills in using the technology may inappropriately influence performance scores.

For instance, an assignment is set to assess the candidate's knowledge in 'embryological development of the cardiovascular system'. Candidates A and B take an online assignment question in 'embryological development of the heart'. Suppose that both candidates have the same level of knowledge in embryological development of the cardiovascular system, but candidate A has superior computer skills than candidate B. Candidate A using her superior computer skills adds a lot of digital images to her assignment. Candidate B does not know how to illustrate her assignment with digital images. Therefore, candidate A gets a higher grade for the assignment than candidate B.

Although the construct under investigation in the example given is ‘knowledge’ (in ‘embryological development of the cardiovascular system’), it has also (unintentionally) assessed the construct of ‘computing skills’; so, the construct of ‘computing skills’ has become a confounding factor, and the assessment has tested an irrelevant or at least unintended construct (Downing, Haldayna 2004). Test developers should therefore ensure that those designing assessments have in the ability to distinguish between the constructs to be assessed.

4.1.2 Real-Life Experiences

Many technology-based assessments attempt to emulate real-life experiences. However, to achieve this goal, the scripts for such assessments need to reflect real-life situations and appropriate level of complexity (Kneebone 2009), particularly since clinical practice is complex and non-linear, and influenced by multiple variables (Epstein Hundert 2002). Minimizing the impact of and interactions between these variables may make the process of assessment easier to handle, but it widens the gap between the assessment activity and the clinical reality that it represents. Conversely, an instructor might design an assessment that contains details or requires actions that unnecessarily increase the complexity of the exercise beyond the learner’s current level of training. This complexity, particularly if it is not germane or intrinsic to the construct being assessed, can increase cognitive load, which in turn can cause performance to suffer. Finally, there is a risk that the designers of technology enhanced assessment activities might select topics that do not reflect situations encountered in typical practice. While in some instances this can be intentional (e.g., assessing performance in response to rare situations), if not germane to the purposes of the assessment then this constitutes an area of potential invalidity (yet another instance of construct-irrelevant variance). Conversely, the omission of important cases would present the problem of construct under-representation.

For instance, an OSCE station has been designed to assess competence in central line placement for a group of 150 trainees using a life-size model of the head, neck, and shoulder. By the time the last student begins, the skin on the model has over 300 holes (most of them localized over a very small area) and the subcutaneous tube representing the jugular vein has begun to leak and so the fluid it once contained has been drained.
The skin holes present a target for students, particularly those later in line that would not be present in life. Simultaneously, those later in line will not be able to assess venous penetration by aspirating blood. Both of these misrepresent reality (and also reflect variation in testing conditions among learners – a source of score invalidity). Also, depending on the learner level and expected competencies, the absence of interpersonal interaction with a live patient could make this procedure overly simplistic (Downing and Haladyna 2004). One should ensure that the measures of assessment are well linked to the practical context rather than what the “simulator" can measure/provide.

4.1.3 Tension between Learner Assessment & Course/Technology Evaluation

Particularly when implementing a new technology, its users might be interested in evaluating the performance of the technology at the same time as it is being used for the assessment of learners' performance. To the degree that these measurements interfere, this could invalidate the inferences about learners.
For instance, a teacher wants to measure team performance in three simulated emergency room scenarios involving hypothetical patients in cardiac arrest. This simulation is a mandatory assessment for all physician trainees, and scores are used as part of the rotation grade. The activity is also part of a research protocol to evaluate the realism of the simulator (mannequin). For the study protocol, participants wear a probe measuring pulse and respiratory rate, and at the end of each simulation exercise trainees indicate their perceptions of a) the authenticity of the mannequin, b) the degree to which they suspended disbelief, and c) the degree to which they feel their actions reflect what they would do in real life. In this simulation assessment, the measures used to evaluate the activity itself are

irrelevant to the construct of team performance. The act of measurement could directly affect their performance (Hawthorne effect), and moreover could paradoxically cause them to pay more attention to the simulation than they otherwise would. This creates a source of invalidity by imposing extraneous demands upon candidates. Identifying and balancing these confounders are critically important when applying study findings in real life applications.

4.1.4 Appropriate Levels of Fidelity

Validity and reliability are linked to the fidelity of representation of clinical contexts and actions within technology enhanced assessment. While greater fidelity enhances the perceived realism of the encounter, it can also increase the complexity and cognitive load associated with the assessment exercise. High-fidelity assessments (e.g. simulation) may be poorly suited for assessing some learning objectives (e.g. knowledge) or may not be well-suited for certain specialties. For example, full immersion simulation with human patient simulators works well for anesthetic teams, but less so when the focus is on surgeons and others doing procedural interventions (Kneebone). Also, higher fidelity usually comes at a price; both monetary cost for the technology itself, and also the cost in instructor time to develop and conduct the assessment. Thus, assessors should target appropriate levels of fidelity for the given assessment task.

For instance, to assess ‘application of knowledge’, students must complete an eight-station objective structured clinical examination (OSCE) comprised of five standardized patients and three stations with high-fidelity mannequin simulators with abnormal exam findings. Each station is observed by a trained faculty member. Knowledge is assessed using a checklist of required actions performed during each station (20%), a global assessment of knowledge (15%), grading of their final write-up using a defined scoring rubric (25%), and accuracy of final diagnosis (40%). The total exam takes about 90 minutes.

This multimodal grading scheme assesses several different aspects of application of knowledge. However, with only eight stations the assessment is likely to suffer from construct under-representation. A MCQ based test of knowledge could present 60-90 questions in the same time period. Although the reliability of each individual MCQ would likely be much lower than the reliability of one OSCE station, the written test more than overcomes this limitation by permitting more questions in the same time, covering a wider variety of topics and situations (Norcini 2002). In contrast, if the construct were "ability to recognize heart sound abnormalities" then a computer-based test could present some features (e.g. heart sounds and on-screen visualization of visible abnormalities) but not others (e.g. abnormalities requiring palpation). If the construct were "interviewing skills" then a standardized patient OSCE would likely be superior. In summary, constant judgment needs to be made to determine level of fidelity required for the given test.

4.2 Affordance and Rapidity of Technological Advances

As technology changes, it frequently outstrips our cultural and professional abilities to understand and incorporate it into our practice. Technology also changes us, often with little awareness of these changes (Gordon Graham). For instance, it is possible to track every interaction in a computer-mediated environment which has led to the summative assessment of contributions to discussion boards; unthinkable and unvalued until it became possible. The gains are fluidity and real time collection of data; however, the potential limitation is the loss of human interaction (Ellaway and Martin 2008) – a valued element in the developmental of the professionals (Tosteson 1979).

Rapid development of technology also poses an ethical challenge of a different dimension. Because of the rapid pace of development, the affordance, or quality of the technology that allows individuals to perform an action, can rapidly outstrip the capacity or interests of those who use it. In addition, rapid changes in technology increase the cost of assessment devices and the quantity and frequency of assessor training, and in some cases could challenge candidates unfamiliar with the new technology. The technology then becomes a barrier to widespread use rather than a facilitator, requiring specialized expertise. For example, efficient use of many high-end high-fidelity simulators (e.g. anesthesia mannequins, trauma scenarios, virtual patients) needs specialized expertise and training that might be outside the scope of many physician educators. This is especially so in assessment where the need for quality scenario development, standardization, and overall quality control is of paramount importance.

As the technology advances it also creates tensions with existing technology and existing structures and amongst users. The agenda from the perspectives of students, teachers, administrators, and device manufacturers are not necessarily synchronous. This mismatch means that evaluation and adoption of cutting-edge technology may be based on factors other than hard evidence - persuasiveness and enthusiasm of inventors, for example, or the rhetorical powers of device manufacturers. These tensions between the agendas of assessors, candidates, developers, institutions, patients, and many others exert powerful effects that are seldom acknowledged (Kneebone 2009). This raises a fundamental issue in identifying the 'driver' for adoption and dissemination of technology in assessment. Experience has shown that the potential euphoria associated with new technology can result in early adoption and then frustrated abandonment (Fenn and Linden 2005). Early adaptors may well produce innovative solutions, but these might be unsustainable, too narrow, insufficiently aligned with learning or quickly superseded by newer technologies. These are illustrated well in clinical skills and procedural learning. As new technologies blossomed in the 1990s (part-task trainers, mannequins of varying fidelity and e-learning tools), health professional programs tended to embrace them by purchasing and then wondering how to use and integrate them – a 'buy first, think later' approach (Bradley & Bligh). Conversely, technology imposed on an institution without adequate planning, infrastructural support, training or a view to sustainability can be detrimental to rather than helpful (Ellaway, Dewhurst et al, 2003).

4.3 Inequity of Resources

In a highly connected world, assessment practices of a given country, which is closely linked to accreditation and licensures, influence other countries (Mullan NEJM; Amin et al 2010). This has significant implications on planning and interpreting assessment results at a global level. Variations in infrastructure, funding, and vendors' presence may create an uneven field with regards to availability and adoption of technology and its deployment in assessment. In an inequitable environment, high resource countries with higher rate of adoption of cutting edge technologies drive technology's innovations, standards, and usages. This raises a question of applicability and transferability of assessment practices and standards among countries and among institutions with varying technological resources.

We conclude by reaffirming that technology should be our tool and not our master. We need to be aware of the scope and capability of current and novel technologies, to collaborate on development and to be open to incorporating them into our educational practice, and to incorporate technology in assessment in a planned, supported and sustainable manner as elegantly expressed as "Learning, Assessment, Technology - In That Order" (Williams D 2009).

5. Future Research

The rapid growth in technology, including the expansion of simulation models and the development of sophisticated test delivery and scoring software, has generated many interesting and important research questions. We outline some of the more relevant research questions concerning the application of technology in assessment. While it is beyond the scope of this paper to outline a specific research agenda related to technology enabled assessment, there are some general issues that must be addressed to ensure that any new technology-based developments, whether used for formative or summative purposes, yield defensible scores and/or decisions.

5.1 Validity: As technology improves, offering more ways to assess candidates, we must continue to be concerned with the validity of the inferences we make based on the assessment scores. In short, just because technological improvements can provide more efficient delivery of assessments, yield higher fidelity test content, and enable rapid scoring and tailored feedback, we still need to ensure the validity of the assessment results. As outlined earlier, one can focus research efforts on investigating potential threats to validity (Wiggins 1993). These types of studies could include looking at the impact of candidate familiarity with the assessment method on ability estimation, timing and pacing issues and, more broadly, the relationship between the fidelity of the assessment and performance. However, the most salient validity issue is establishing the relationship between performance on the assessment and, for healthcare practitioners, the performance with 'real' patients. Longitudinally, we need to determine whether advances in assessment technologies lead to better patient outcomes {refs} or other benefits such as economics or efficiency. Also, given the prevailing literature on impact of assessment on learning (Larsen 2009; Galvagno and Segal 2009), and the enhancements in the fidelity of various assessments {ref}, comparative research aimed at quantifying the educational impact of new assessment formats is certainly called for.

5.2 Reliability: Technological advances, including the development of computer-based delivery of test content and the evolution of part-task trainers, OSCEs, and electromechanical mannequins, have allowed for the construction of many new and different types of assessment processes. Like all assessments, however, the sources of measurement error need to be investigated and quantified. For OSCEs, especially those involving standardized patients, computer-based training of patient actors can enhance the fidelity of their portrayal and minimize scoring errors {refs}, thus yielding more reliable estimates. Additional research concentrating on the application of technology for the training of those involved in healthcare-related performance assessments, including raters, is needed. Finally, with the introduction of physical and onscreen simulators into the assessment domain, test developers have been challenged with the construction of new scoring rubrics (Margolis et al, 2004). For some assessments, such as those involving procedural skills, the evaluation tools will be case-specific, potentially limiting the generalizability of the scores. For others, including simulation scenarios keyed to measuring more generic skills such as teamwork and ethical behavior, the evaluation tools and scoring criteria can be difficult to interpret, even for experts, resulting in evaluations that can be highly subject to rater effects. As advances in technology broaden the assessment domain, it is important that research be conducted to determine the specific sources of measurement error in the scores.

5.3 Other Research Areas: While research concerning the validity and reliability of assessment scores is paramount, technological advancements in test delivery, test construction and simulation for assessment also provide other opportunities for targeted studies. One of the most important research areas rests in ascertaining the comparative efficiency of competing testing approaches. While higher fidelity assessment models may be perceived to be more effective in measuring educational outcomes, they are costly, can be logistically complex, and may not yield appreciably better measures of ability {refs}. Likewise, if technology is used to deliver formative assessments research is needed to best align educational models with the learning needs of the participants. These outcome measures can eventually be used validate prior assessment scores.

Consensus Statement and Recommendations

General recommendations

1. Institutional leaders, teachers, and other stakeholders should understand and follow general principles of quality assessment when using technology-enabled assessment.
2. Educators and leaders should not feel compelled to use technology if it does not extend their capabilities. Indeed they should use caution in employing technologies that do not serve a demonstrable purpose or otherwise enable or extend current capabilities. Educational need, resource efficiency, and validity should be the primary deciding factors in choosing the right technology for enabling or enhancing assessment.

Recommendations for Institution Leaders and Policy Makers

In institutions that elect to employ technology-enabled assessment:

1. Leaders should ensure effective integration of technology-enabled assessment with other assessment activities.
2. There should be an integrated institutional responsibility for the overall planning, integration, and implementation of technology-enabled assessment in the curriculum.
3. There should be a coherent institutional strategy to maximize the efficiency and return on investment of both technical and non-technical resources in all aspects of assessment.
4. Curriculum and Assessment Committees should include member(s) with expertise in technology-enabled assessment to facilitate the appropriate planning, integration, and implementation of technology-enabled assessment.
5. Institutional leaders should provide for appropriate faculty and student development in using technology-enabled assessment.

Recommendations for Assessment Committee and Planners

1. The application and purpose of technology-enabled assessment should be made clear to all stakeholders, capitalizing on specific advantages and complimenting existing methods.
2. Assessment must take into account local technical contexts prevailing in clinical and educational environments and should make appropriate use of these technologies in designing the assessment.
3. Evaluators should consider potential monetary and non-monetary conflicts of interest between device manufacturers, promoters, administrators, and other users of technology.

Recommendations for Individual Teachers and Developers

1. Design of the assessment environment should be enabling, accessible and appropriate to the needs of the assessment and of the participants.
2. Prior to participating in a technology-enabled assessment, candidates should have adequate instruction in and practice opportunities with the assessment technology employed.
3. Individual teachers and developers should take a proactive approach towards personal and professional development in the use of technology in assessment.

Future Research

1. Establish continuity: Researchers should study the relationship between performance of assessment in the simulated environment and performance in real-life.
2. Clarify roles: Researchers should study the application of different technologies to specific contexts, to inform decisions regarding which technology to use for what learning context.
3. Establish scoring procedures: Researchers should develop scoring methods that automate the collection, integration, and analysis of the vast and often novel information available through technology-enabled assessment.

4. Connections: Researchers should establish better and more robust links between workplace systems and patient outcome data and their use for assessment, including - but not limited to - portfolios, logbooks and chart reviews.

References:

1. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association; 1999.
2. Amin Z, Chong YS, Khoo HE. Practical Guide to Medical Student Assessment. World Scientific Publishing Company. 2006. Singapore.
3. Amin Z, Burdick WP, Supe A, Singh T. Relevance of Flexner Report to Contemporary Medical Education in Asia. Academic Medicine. 2010. 85 (2): 333-9.
4. Bradley P. The history of simulation in medical education and possible future directions. Medical Education. 2006. 40 (3): 254-262.
5. Black SA; Nestel DF, Horrocks EJ, Harrison RH; Jones N, Wetzel CM, Wolfe JHN; Kneebone, RL. Evaluation of a Framework for Case Development and Simulated Patient Training for Complex Procedures. Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare. 2006. 1(2): 66-71.
6. Black SA, Nestel DF, Kneebone RL, Wolfe JHN. Assessment of surgical competence at carotid endarterectomy under local anaesthesia in a simulated operating theatre. British Journal of Surgery. 2010. 97(4): 511-516.
7. Bennett RE. Using New Technology to Improve Assessment. Educational Measurement: Issues and Practice. Volume 18 Issue 3, Pages 5 – 12. 25 Oct 2005. Online at http://www3.interscience.wiley.com/journal/119079459/abstract?CRETRY=1&SRETRY=0
8. Biggs J & Tang C. Teaching for Quality Learning at University. 3rd edition, Open University Press, Maidenhead. 2009.
9. Boulet JR, Murray D, Kras J, Woodhouse J, McAllister J, Ziv A. Reliability and Validity of a Simulation-based Acute Care Skills Assessment for Medical Students and Residents. Anesthesiology. 2003. 99: 1270-80.
10. Boulet JR. Summative Assessment in Medicine: The Promise of Simulation for High-stakes Evaluation. Academic Emergency Medicine. 2008. 15(11): 1017 – 1024.
11. Bradley P, Bligh J. Clinical Skills Centres: where are we going? Med Educ. 2005 Jul; 39(7):649-50.
12. California Health Care Foundation. How Smart Phone is Changing Health Care for Consumers and Providers. April 2010. Online at http://www.chcf.org/publications/2010/04/how-smartphones-are-changing-health-care-for-consumers-and-providers. Last accessed June 23, 2010.
13. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull. 1959;56:81-105.
14. Cashion J, Palmieri P. 'The secret is the teacher': the learner's view of online learning. Web address: http://www.ncver.edu.au/research/proj/nr0F03e.htm. Last accessed June 23, 2010.
15. Carroll JD. Messenger JC. Medical simulation: the new tool for training and skill assessment Persp Biol & Med. 2008.
16. Cook DA, Beckman TJ. Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application. Am J Med. 2006;119:166.e7-16.
17. Cooper JB. A brief history of the development of mannequin simulators for clinical education and training. Postgrad Med J. 2008. 84:563-570.
18. Chou P, Chen, W. From Portfolio to E-Portfolio: Past, Present, and Future. In K. McFerrin et al. (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2008.* 22-27. 2008. Chesapeake, VA: AACE.
19. Dev PP. Youngblood, et al. (2007). "Virtual Worlds and Team Training." Anesthesiology Clinics 25: pp321-336.
20. Dillon GF, Boulet JR, Hawkins RE, Swanson DB. Simulation in the United States Medical Licensing Examination ™ (USMLE™). Quality and Safety in Health Care. 2004. 13: 41-45.
21. Donnelly RCAL, Gorman M P (1999) Planning and Developing an Interactive Computerised Tutorial for Learning in Higher Education *Teaching in Higher Education* 4, 3, 397-410
22. Downing SM. Validity: on the meaningful interpretation of assessment data. Med Educ. 2003;37:830-837.
23. Downing SM. Reliability: on the reproducibility of assessment data. Med Educ. 2004;38:1006-1012.
24. Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. Med Educ. 2004;38:327-333.
25. Haines C (2004) *Assessing Students' Written Work. Marking Essays and reports.* Routledge Falmer London, UK.

26. Ellaway R, D. Dewhurst, et al. (2003). "Managing and supporting medical education with a virtual learning environment - the Edinburgh Electronic Medical Curriculum." Medical Teacher 25(4): 372-380.
27. Ellaway R, Martin, R. (2008). "What's Mine Is Yours - Open Source as a New Paradigm for Sustainable Healthcare Education." Medical Teacher 30(2): pp175-179.
28. Ellaway, R. and Masters, K. (2008). "AMEE Guide 32: e-Learning in medical education Part 1: Learning, teaching and assessment." Medical Teacher 30(5): pp455-473.
29. Ellaway RH, Kneebone R, Lachapelle K, Topps D. Practica continua: Connecting and combining simulation modalities for integrated teaching, learning and assessment. Medical Teacher. Volume 31, Issue 8 August 2009 , pages 725 – 731.
30. Epstein RM, Hundert EM. Defining and Assessing Clinical Competence. JAMA. 2002. 387: 226-235.
31. Epstein RM. Assessment in Medical Education. NEJM 2007. 356:387-396.
32. Fenn, J. and A. Linden. (2005). "Gartner's Hype Cycle Special Report for 2005." Online at http://www.gartner.com/DisplayDocument?doc_cd=130115 last accessed 12 June 2010.
33. Foster SL, Cone JD. Validity Issues in Clinical Assessment. Psychological Assessment. 1995;7:248-260.
34. Galvagno SM & Segal BS. Critical Action Procedures Testing: A Novel Method for Test Enhanced Learning. Medical Education 2009: 43: 1182-1187.
35. Gordon JA, Wilkerson WM, Shaffer, DW, Armstrong, EG. “Practicing" Medicine without Risk: Students' and Educators' Responses to High-fidelity Patient Simulation. Academic Medicine. 2001. 76 (5): 469-472.
36. Farfel A, Afek A, Hardoff D, Ziv A. The Effect of a Simulated-Patient-Based Educational Program on Medical Encounters' Quality at Military Recruitment Centers. Accepted for publication in IMAJ; 2010.
37. Fried MP, Satava R, Weghorst S, Glallagher AG, Sasaki C, Ross D et al. Identifying and Reducing Errors with Surgical Simulation. Quality and Safety in Health Care. 2004. 13: 19-26.
38. Gesundheit, NP. Brutlag, et al. (2009). "The Use of Virtual Patients to Assess the Clinical Skills and Reasoning of Medical Students: Initial Insights on Student Acceptance." Medical Teacher 31(8): pp739-742.
39. Graham, G. (1999). The Internet: A Philosophical Inquiry, Routledge. London UK.
40. Greenhalgh, T. (2001). Computer assisted learning in undergraduate medical education, British Medical Journal. 322 , 40-44.
41. Hays R, Singer M. Simulation fidelity in training system design: Bridging the gap between reality and training. Springer-Verlag. 1989.
42. Hodges B. Medical education and the maintenance of incompetence. Med Teach. 2006;28:690-696.
43. Issenberg SB, McGaghie WC, Hart IR, Mayer JW, Felner JM, Petrusa ER, Waugh RA; Donald D. Brown, MD; Robert R. Safford, MD, PhD; Gessner IH, MD; Gordon DL, MD; Ewy GA. Simulation Technology for Health Care Professional Skills Training and Assessment. JAMA. 1999; 282:861-866.
44. Issenberg SB, McGaghie WC, Petrusa ER, Gordon DL, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teacher 2005;27(1):10–28.
45. Kane MT. Validation. In: Brennan RL, editor. Educational Measurement. 4th ed. Westport: Praeger; 2006. p. 17-64.
46. Keenan, C. R., H. H. Nguyen, et al. (2006). "Electronic Medical Records and Their Impact on Resident and Medical Student Education." Academic Psychiatry 30(6): pp522-527.
47. Kneebone R. Practice, Rehearsal, and Performance An Approach for Simulation-Based Surgical and Procedure Training. JAMA. 2009;302(12):1336-1338
48. Kneebone R. Perspective: Simulation and Transformational Change: The Paradox of Expertise. Academic Medicine. 2009. 84 (7): 954-957.
49. Kneebone R, Kidd J, Nestel D, Asvall S, Paraskeva P, Darzi A. An innovative model for teaching and learning clinical procedures. Medical Education. 2002. (36): 628-634.
50. Kneebone R, Arora S, King D, Sevdalis N, Kassab E, Aggarwal R et al. Distributed simulation - Accessible immersive training. Med Teach. 2010; 32(1):65-70.
51. Kress G. Multimodality: a social semiotic approach to contemporary communication. London: Routledge, 2010.
52. Larsen PD, Butler A, Roediger III HL. Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. Medical Education 2009: 43: 1174–1181.

53. Margolis, M., B. Clauser, et al. (2004). "Scoring the Computer-Based Case Simulation Component of USMLE Step 3: A Comparison of Preoperational and Operational Data." Academic Medicine 79(10): ppS62-S64.
54. Miller GE. The Assessment of Clinical Skills/Competence/Performance. Academic Medicine. 1990. 65(9): S63-S67.
55. Mullan F. The Matrices of physician brain drain. NEJM. 2005: 353: 1810-1818.
56. Messick S. Validity. In: Linn RL, editor. Educational Measurement, 3rd Ed. New York: American Council on Education and Macmillan; 1989.
57. Norcini JJ. The Death of Long Case. BMJ. 2002. 324: 408-409.
58. Norcini JJ, McKinley DW. Assessment Methods in Medical Education. Teaching and Teacher Education. 2007. 23: 239-250.
59. Premkumar K, Coupal C. Rules of engagement-12 tips for successful use of “clickers” in the classroom. Med Teach. 2008.30(2): 146 – 149.
60. Reiser, S. J. (2009). Technological Medicine: the Changing World of Doctors and Patients. New York, Cambridge University Press.
61. Scalese RJ, Obeso VT, Issenberg SB. Simulation Technology for Skills Training and Competency Assessment in Medical Education. JGIM. 2007. 23(Suppl 1):46–9
62. Scarborough H and Corbett, JM. (1992). Technology and organization: power, meaning and design UK, Routledge.
63. Schön, D. A. (1987). Educating the Reflective Practitioner. USA, Jossey-Bass.
64. Schuwirth LW, van der Vleuten CP. A plea for new psychometric models in educational assessment. Med Educ. 2006; 40: 296-300.
65. Schuwirth LWT, van der Vleuten CPM, De Kock CA, Peperkamp AGW, Donkers HHLM. Computerized Case-based Testing: A Modern Method to Assess Clinical Decision Making. Medical Teacher. 1996. 18(4):294-299.
66. Shavit I, Keidan I, Hoffmann Y, Mishuk L, Rubin O, Ziv A, Steiner I. Enhancing Patient Safety During Pediatric Sedation: The Impact of Simulation-based Training of Non-Anesthesiologists. Arch Pediatrs Adolesc Med; 161(8):740-743, 2007.
67. Shachak A, Hadas-Dayagi M, Ziv A, Reis S. Primary care physicians’ use of an electronic medical record system: a cognitive task analysis. J Gen Intern Med; 24(3):341-8, 2009
68. Sica GT, Barron DM, Blum R, Frenna TH, Raemer DB. Computerized Realistic Simulation: A Teaching Module for Crisis Management in Radiology. American J Radiology. 1999. 172: 301-4.
69. Innovate Simulations for Assessing Professional Competence: From Paper-and-Pencil to Virtual Reality. Tekian A, McGuire CH, McGaghie WC. Chicago. Dept of Medical Education, University of Illinois at Chicago. 1999.
70. Tosteson D. Learning in Medicine. NEJM. 1979: 301 (13): 690-4.
71. Tucker, B. Beyond the Bubble Technology and the Future of Student Assessment. Education Sector Report. Feb 2009. Education Sector. Washington DC. USA.
72. van der Vleuten CPM, The assessment of professional competence: Developments, research and practical implications. Adv Health Sci Educ 1996. 41-67.
73. William D. “Learning, assessment, technology: in that order.” Plenary presentation at the Association of Medical Education in Europe Conference. September, 2009, Malaga, Spain.
74. Wireless Health Care When your carpet calls your doctor. Economist. 8 April 2010. New York. Web address: http://www.economist.com/business-finance/displaystory.cfm?story_id=15868133&source=hptextfeature
75. Wiggins G. Assessing Student Performance exploring the purpose and limits of testing. San Francisco: Jossey- Bass, 1993.
76. Wong SH, Ng KF, Chen PP. The Application of Clinical Simulation in Crisis Management Training. Hong Kong Med J. 2002. 8: 131-5.
77. Ziv A, Wolpe PR, Small SD, Glick S. Simulation-Based Medical Education: An Ethical Imperative. Acad. Med. 2003;78:783–788.

Annex A: Definition of Terms

The lack of reliable and consistent terms that are able to incorporate, typify and model emerging techniques is a barrier to a grounded empirical approach. The first step, therefore, is to define what we mean by technology along with a number of other key terms that will be used and reused in this document. This has proved to be particularly important in creating a consensus statement where the meanings of these terms vary between the participants' loci of practice. We therefore define the following:

Technology: Technology is a hugely complex construct and includes both 'machines' and 'processes'. National Center for Biotechnology Information (NCBI) defines technology as "the application of scientific knowledge to practical purposes in any field. It includes methods, techniques, and instrumentation." (MeSH, NCBI). Whereas, educational technology is defined as "systemic identification, development, organization or utilization of education resources and the management of these processes." (Thesaurus of ERIC Descriptors; Dec 1993, p 132). In this consensus statement, we define technology as "any tool/method innovated through the advancement of science that would assist and enhance traditional assessment and may include machines and tools as well as processes or a combination of both." Thus, in the broadest sense, technology includes machines such as simulators, human beings in the role of simulated patients, and to more process-driven innovations such as Electronic Medical Records (EMR) and digital archives. It also includes the methods and processes used for working with data and other information systems and artifacts associated with assessment and healthcare professional education as a whole.

e-Assessment: those assessment processes mediated by information and communication technologies. (We have not used the term in the main body,)

Information and communication technologies: those technologies involved with managing and manipulating information and mediating communication. ICTs include computers (desktop, laptop, notebook), mobile/smart phones, peripherals (screens, mice, printers), networks (the Internet) and the software that allows all of these machines to operate.

Informatics: all of the principles, models and phenomena associated with the use of ICTs. Informatics is concerned with how ICTs are used, the impact they have and the underlying patterns and practices that define them.

Simulation and simulator: in the broadest sense simulation is the process of recreating an environment, situation or context. Simulators, simplistically speaking, are machines and devices that include mannequins, computers, and models. It also encompasses human simulators as standalone proxy participants (not necessarily patients) and hybrid models (human plus device). In an ideal situation, simulators and simulation should exist together balancing the authenticity of the clinical environment and educational priorities.

Fidelity: Fidelity refers to the degree to which a model or simulation reproduces the state and behaviour of a real world object, feature or condition. Fidelity is therefore a measure of the realism of a model or simulation (Fidelity Implementation Study Group Report). Simulation fidelity has also been described in the past as 'degree of similarity' (Hays 1998).

Annex B: Contemporary Applications and Illustrations

There are many applications of technology in assessment in healthcare education and training. The usage can be broadly divided into two categories: assessment tools and technologies that deal with assessment of candidates and the others that deal with management of assessment processes. Essentially, all the illustrations used for assessment of candidates can be used for formative purposes and should have elements of feedback to improve learning.

We present the following illustrations as a representative (but not exhaustive) range of contemporary applications of technology-based assessment. Each illustration is comprised of the following elements:

- Players: those actors and other agents involved in the activity
- Technologies used: all systems, tools and technical processes involved in the activity
- Strategies: the educational/assessment strategies addressed involved in the activity
- Typical description: a stepwise narrative description of the activity
- Data used: what data is used and how it is generated
- Variations/Innovations: instances where the description of events might be changed to represent a wider range of forms of use

Assessment Tools and Technologies

Online knowledge based exams	
Players	Candidates, invigilators
Technologies Used	Commercial web-based e-assessment tool such as ‘QuestionMark Perception’ and other homegrown systems The Exam is served from and data is written to a secure server. Only data from the server is used in marking.
Strategies	Exam capable of testing multiple domains and some hierarchies of learning
Typical Description	An exam is presented using a commercial e-assessment platform combining a number of different formats: • MCQ – these involve selecting either best option or true/false for each option presented • type in text questions – these involve a mixture of short text answers to a question and filling in missing words in a paragraph of text • matching questions – these involve matching a series of items using a combination of dropdown text and ‘drag and drop’ items. • image interaction – these involve the candidate clicking on a an image to identify a particular structure indicated in the question The exam involves all of the learners sitting in a single room together with each candidate sitting at a computer terminal with screen, keyboard and mouse. Each candidate logs in to the exam system using his or her secure institutional credentials. This automatically launches a locked down web browser that only lets the candidate access the exam system. The exam starts with the master invigilator clicking a release button and the exam is closed with the master invigilator clicking a close button. Candidates can complete the questions in any order, they can change their answers right up until the end of the exam and they can flag those items they wish to return to for further consideration.
Data used	Data consists of item-response pairs – digital only
Variations/Innovations	• Other question types are used • Distributed system • No master start and stop control • Candidates login using credentials handed to them expressly for the current exam • Once complete a ‘page’ or section of questions cannot be further edited • Dedicated client tool rather than web browser used • Feedback on the given answer • Progress tests delivered across a programme

Online Examination of Applied Clinical Knowledge, Reasoning: Cases, Virtual patients (VP)	
Players	Candidates, invigilators
Technologies Used	Virtual patient case engines: • Branching designs: OpenLabyrinth, vpSim • Schema designs: WebSP, CASUS Exam is served from and data is written to a secure server. Only data from the server is used in marking.
Strategies	Exam tests decision making, synthesis of multiple knowledge domains and patient management
Typical Description	An exam is presented as a series of case scenarios and vignettes. The shorter vignettes are variations on the 'key feature question' format each of which may include extended matching or type in text questions and responses. The case scenarios are virtual patient activities requiring the candidate to negotiate a series of decisions in diagnosing and treating a patient. Candidates may only tackle each vignette or case once and they may not go backwards within a question. They can tackle the vignettes and cases in any order they chose.
Data used	Data consists of item-response pairs for the vignettes and a series of decision pathways or item requests for the virtual patient cases – digital only
Variations/Innovations	• Drop-down menu for options • Mixed media with images, videos, audio recordings • Ability to adapt to local setting

Examination of Applied Sciences: Simulation	
Players	Candidates, invigilators (in-room and operators)
Technologies Used	High fidelity mannequins with a computer control interface e.g. Laerdal SimMan, METIMan
Strategies	Exam tests decision making, clinical skills, teamwork
Typical Description	An exam is presented as a simulation exercise with one candidate being tested at a time (or team assessment). The scenario starts as soon as the candidate enters the exam room and lasts for a pre-defined duration. The candidate is briefed on their role and expected outcomes and introduced to the 'patient' (represented by a high fidelity mannequin). The candidate is expected to take on the given role and manage or respond to the situation, the patient and other members of the team.
Data used	Hybrid: the digital log files generated by the mannequin plus direct observations
Variations/Innovations	• Assessment of multiple learners at one time (team) • Assessment of complex drills (e.g., resuscitation) • Assessment of rare or life-threatening events

Clinical Skills Examination: Simulators	
Players	Candidates, invigilators
Technologies Used	Human actors taking the patient role, simulators or mannequins with or without processing capabilities; e.g., Harvey, various task trainers.
Strategies	Exam tests physical examination and practical skills, decision making,

	and patient management
Typical Description	Each candidate is given a defined task involving the demonstration of one or more clinical skills, along with appropriate knowledge and attitudes. The candidate performs the assigned clinical task on the simulated patient/mannequin observed by the examiner who marks off aspects of performance on a checklist. This may be followed by brief discussion between examiner and candidates. There may be interaction between ‘patient’ and candidates.
Data used	Direct observation using pre-defined checklist
Variations/Innovations	• Hybrid with part task trainer coupled with real human being for enhanced reality (e.g., a breast examination model is attached to a real human being) • ‘Patient’ marks the candidates • feedback on performance (from examiner, ‘patient’ actor)

Electronic OSCE stations	
Players	Candidate, assessors
Technologies Used	Training version of EMR, PACS, electronic prescription system
Strategies	Ability to use medical informatics and patient management systems
Typical Description	Each candidate is given a defined task involving taking a patient history and entering it into the EMR and accessing relevant radiology from the PACS. The candidate performs the assigned clinical task using the computer terminal provided and is observed by the examiner who marks off aspects of performance on a checklist. This may be followed by brief discussion between examiner and candidates.
Data used	Predetermined checklists paired with audit of data entry and computer logs
Variations/Innovations	▪ Linked with real or simulated patient cases

Multidimensional Record of Candidate’s Performance: e-Portfolio	
Players	Candidate
Technologies Used	PC Diary Electronic portfolio system (many available)
Strategies	Assess professional development over time and across multiple integrated dimensions
Typical Description	Candidate builds up portfolio over a period of time with a mind to demonstrate: a) ‘all of me’ – the negative and weak aspects of the individual as well as their strengths b) ‘best of me’ – the positive and optimal aspects only Entries may be elective or mandatory Entries may be structured, for instance, through a pro forma reflective template or unstructured through a blog or other posting mechanism Access by others to entries is set so that peers and/or tutors can view and perhaps comment on different entries Entries may be tagged by the learner to represent certain professional competencies or different kinds of events The portfolio presents entries as a structured collection for self- and summative assessment purposes
Data used	Any data inserted – textual accounts, images, videos, links to other materials, annotations etc. Pro forma reflective exercises Comments by peers and tutors
Variations/Innovations	▪ Log books which records clinical encounters, procedures performed ▪ Reflective journals

	▪ Free-flowing structure and form determined by the candidate

Record of Candidate Performance during Group Project	
Players	Candidates work as teams
Technologies Used	Wikis, blogs, other social-networking tools
Strategies	Assess research skills, ability to work within as a group
Description	A group of learners is assigned a project. They work together to create a website outlining their studies using a wiki provided by the school. Contributions, change histories and edits are tracked and the tutor invited to comment as the site comes together. Finished site is presented formally to class and to tutors and grading is done via an aggregate of ratings from all tutors of all groups.
Data used	Raters complete rating form based on observations and predetermined outcome criteria
Variations/Innovations	▪ Peer assessment ▪ Inter-professional assessment ▪ Assessment across multiple institutions

Assessment Management Technology

Logistics – Authoring	
Players	Faculty authors, exam coordinators
Technologies Used	Authoring system (e.g. Respondus), item banking system, delivery system
Strategies	Streamlining, tracking and analyzing item authoring process
Description	An exam is being assembled (either electronic or paper-based) Three groups of authors are preparing questions for the exam: a) The first set of authors are creating questions from scratch b) The second set of authors are adapting existing questions from an item bank using past performance metrics for each item to assess its utility in the current exam context c) The third set of authors import questions from an external source into the item bank and then into the exam question pool The exam coordinator assembles exam papers by selecting and sequencing items based on topics/objectives covered, flags for not pairing certain sets of questions and overall time available and types of questions required. Once assembled the exam is tested and edited, if necessary, before delivery.
Data used	Usage information Items statistics (e.g., difficulty and discriminatory index)
Variations/Innovations	▪ Mixed media (e.g., video, audios, graphics) ▪ Students contributing questions as authors (e.g. 'Peerwise')

Logistics: Marking, Analysis and Reporting	
Players	Faculty assessors, administrators
Technologies Used	Plagiarism detection tool such as 'Turnitin'
Strategies	Assistance in detecting plagiarism
Typical Description	Candidates prepare written project and/or case reports for summative assessment and submit them for marking through their program learning management system. Once submitted the tutor submits all of the reports to the plagiarism detection service for analysis. The following day the tutor receives a report for each piece of coursework submitted comparing the text submitted with a) the other texts submitted, b) literature on the web, c) other coursework submitted through the service by other tutors and institutions and previous work

	from the same tutor. The reports identify the extent of matching between texts and the sources. The tutor assesses the extent and severity of plagiarism and takes the appropriate action.
Data used	Electronic version of assignment in text format
Variations/Innovations	▪ Students are asked to run an electronic version of their assignment through the application and then to submit a report with the assignment.

Logistics: Aggregate Assessment Access and Tracking	
Players	Faculty assessors, administrators
Technologies Used	Handheld devices; computers in the learning environment
Strategies	Collect on-the-go data about candidate's performance (e.g. electronic mini-CEX, DOPS, voice recording or log-books)
Description	Program-wide online assessment tracking system collects data from multiple assessors about candidate's performance from multiple encounters and multiple time-points. Software engine processes multiple inputs and forms to create aggregate performance metrics. Performance metrics displayed on a dashboard accessible to the learner, their tutors and the assessment office – these indicate both the candidate's progress and the qualitative and quantitative aspects of their clinical experience.
Data used	Variable: checklists, patient records, conditions or diseases seen
Variations/Innovations	▪ Learner tracks her own progress ▪ Faculty identifies problem learners; rewards excellent performers ▪ Log book to record clinical experience