Original Article

Next generation of procedural skills curriculum development: Proficiency-based progression

Richard M. Satava[1], Anthony G. Gallagher[2]

[1]Department of Surgery, University of Washington Medical Center, Seattle, Washington, USA, [2]ASSERT Center, School of Medicine, University College Cork, Cork, Ireland

ABSTRACT

Background: Until recently training, assessment and certification of technical procedural skills (surgical and other interventional procedures) has been a matter of subjective evaluation by a faculty or senior peer. A few new technical skills courses have begun quantifying surgical performance, however none have required benchmark scores that must be met to guarantee proficiency. Based on nearly 100 years of technical skills simulation in other high-risk sectors (aviation, military, nuclear industry, etc.), a new, comprehensive procedural skills curriculum development process, entitled 'full life-cycle curriculum development' and which uses the principles of proficiency-based progression, has been developed and implemented for the Fundamentals of Robotic Surgery (FRS) curriculum – a simulation course for technical skills with quantitative metrics.

Methodology: Four consensus conferences by Delphi method, with official representatives of 14 surgical specialties that use robotic surgery was conducted to develop Outcomes Measures/metrics, Didiactic Curriculum, Simulation Content for a novel simulator, and Validation Trial. Proficiency-based progression, based upon expert performance benchmarks, was used to establish the scores which the novices must complete in order to pass.

Results: An initial pilot study for validation of this new curriculum process has been completed to determine rough-order-magnitude of parameters and usability and practicality, as well as a preliminary evaluation of suitability as an educational course. A final full validation study is in progress to confirm the initial results of both the validity and educational value of this course. A validated course for gynecologic robotic surgery has been completed and validated using the same templates as the FRS.

Conclusion: The FRS use a new process (full life-cycle curriculum development with proficiency-based progression) which can be used in order to develop any quantitative procedural curriculum, through generic templates that have been developed. Such an approach will dramatically decrease the cost, time and effort to develop a new specific curriculum, while producing uniformity in approach, inter-operability among different curricula and consistency in objective assessment. This process is currently online, open source and freely available, to encourage the adoption of a scholarly and rigorous approach to curriculum development which is flexible enough to be adopted and adapted to most technical skills curriculum needs.

Keywords: Curriculum development, proficiency-based progression, quantitative assessment, simulation

INTRODUCTION

This is the era of 'evidence-based medicine', in which the practice of medicine is evolving to a system which requires not only 'transparency' in the quality and background of publications, but also one which requires the rigorous use of the scientific method to guarantee the quality of evidence that assures that the proposed clinical practice is both valid and of the highest quality, especially in terms of patient safety. One essential component that has been nearly completely ignored is ensuring that the quality of the educational process

Address for correspondence:
Prof. Richard M. Satava, Department of Surgery, University of Washington Medical Center, Seattle, Washington, USA.
E-mail: rsatava@u.washington.edu

Access this article online	
Quick Response Code:	Website: www.thejhs.org
	DOI: 10.4103/1658-600X.166497

How to cite this article: Satava RM, Gallagher AG. Next generation of procedural skills curriculum development: Proficiency-based progression. J Health Spec 2015;3:198-205.

is of to the same high standards. The following is the first iteration (and validation) of this newly developed process, the 'full life-cycle curriculum development' process. Much of the process is based upon Gallagher's 'proficiency-based progression' (PBP) method,[1] and which is fully described in the book 'Fundamentals of Surgical Simulation: Principles and Practice: Improving Medical Outcomes – Zero Tolerance'.[2]

Quantifiable training curricula (courses) with an objective assessment of performance began being developed after the end of World War I and culminated in the first Link Flight Simulator in 1929. In the intervening years, the fidelity of the simulation had become extraordinary, with a concomitant improvement in the training and assessment methodologies, largely due to the progress in behavioural psychology, psychometrics and related educational systems. In addition, similar developments occurred in parallel high-risk professions, such as other forms of transportation, of nuclear materials and more recently space flight. However, for the medical profession, it was not until the late 1980s that Gaba and DeAnda[3] began the introduction of manikin-based simulation for the equivalent of aviation's 'crew-resource management' which has evolved into today's 'team training and communication (TT and C) skills', or more – broadly under the umbrella of non-technical procedural skills (to distinguish these skills from the technical psychomotor skills). This huge advance in medical training is still evolving, since the current versions (including the standardised TeamSTEPPS[4] which was also partially funded by the Department of Defense) remains an educational experience rather than a quantified performance assessment. While this methodology of evaluating team performance has developed excellent debriefing processes following a training scenario (including some checklists), at the completion of the debriefing, the learner(s) clearly have learned about their errors (in patient safety), however they are not required to repeat the training scenarios until no errors occur, nor are there specific proficiency benchmarks which must be achieved. In essence, they have learned what they have done wrong but are not required to meet standards which have been set to prove that they have learned the material, nor are they truly assessed as having 'passed' or 'failed'. This leaves the privileging (and credentialing) authorities, such as the hospital committees or the board examiners, with no quantitative measure of whether the learners are actually competent. This opportunity to quantify non-technical skills is also proposed herein.

In addition, the psychomotor skills were first quantified in a virtual reality (VR) surgical simulator in 1993.[5] There subsequently was a rapid development of both physical and computer-based (internet, augmented reality, VR, etc.,) simulators of skills and a number of simple procedures. In the beginning, such simulations were created by companies that enlisted an enthusiastic 'expert', who guided the engineers on 'how I do a specific procedure', using easily computed measures (such as time, ambidexterity, etc.). Unfortunately, the results were not acceptable to the accreditation authorities (such as the surgical 'boards') because the outcome measures that needed to be measured were not based upon patient safety (i.e., avoidance of any errors) and not on how fast a procedure could be performed – there is nothing worse than a surgeon who is extremely fast, but makes many errors that imperil patient safety. It was not until the Objective Structured Assessment of Surgical Skills in 2000,[6] validation of a simulator-based VR curriculum[7] and Fundamentals of Laparoscopic Surgery high-stakes testing (HST)[8] that quantitative measures of performance were begun. A great improvement in assessing technical skills, and focusing upon avoidance of errors, a final 'score' was arbitrarily set. Unlike the other high-stakes professions mentioned above, the medical learner was not required to meet the highest standards (as determined by expert performance), but an arbitrarily lower score, such as 75% of the information/skills presented. Imagine a pilot taking off and landing who has demonstrated that they can successfully accomplish both tasks without error (crash) 75% of the time. Thus, a process to determine such 'expert benchmarks', which a learner must achieve before progressing to the next level of training is proposed as a critical component in the full life-cycle development of curriculum in a PBP manner.

In curriculum development, training and assessment are two sides of the same coin[9] – it is essential during training with deliberate practice to provide feedback as soon as an error occurs – referred to as 'formative feedback' as opposed to 'summative feedback', which is feedback that is given at the completion of a trial/task as a summation report of all the errors that occurred. As will be seen below, the teaching of how to make a mistake (error) is a critical part of the educational process and a unique property of simulation-based training – which is the opportunity to make mistakes: In essence encouragement and 'permission to fail', without consequence to patient safety. By clearly teaching what are the common mistakes, how to avoid making such mistakes, how to recognize when a mistake is made and how to remediate (repair) a mistake to avoid serious consequences are all very powerful educational tools, and which are no longer acceptable methods of training on patients in the operating room.

In 2010, the ad hoc committee Alliance of Surgical Specialties in Education and Training (ASSET) met to develop templates which would facilitate and standardise the approach to developing curricula. The curriculum template [Table 1] was described by Zevin *et al*.[10]

However, developing the didactic, psychomotor skills and team training curriculum is not sufficient; it is essential to validate the curriculum (course). There must be validation of both quality of the training and assessment tool(s), as well as validation that training to PBP in a simulation environment transfers to proficiency in a comparable real-world (clinical) situation, referred to as transfer of training (ToT). For validating a course, there are the standard 5 'validities', with simplified explanations as to their purpose:

- **Face:** Looks like what it is supposed to simulate (on face value)
- **Content:** Teaches and measures the correct knowledge and skills
- **Concurrent:** Is as good as, or better than, the current curriculum
- **Construct:** Design (construction) actually shows experts are better than novices
- **Predictive:** Predicts that learners, who do well on simulation, do well on patients.

Table 1: ASSET template of curriculum development template

Needs assessment (requirements document)	
Inventory and gap analysis (see separate pre-course template)	
Goals and objectives	
Outcomes measures	
Includes societies, boards, ACGME, etc., input to determine outcomes measures	
Benchmarked to criterion measures by experienced/expert surgeons/physicians	
Criterion-unambiguously defined with quantitative measures	
Didactic component (cognitive skills)	
Pre-didactic suggested readings	
Anatomy or laboratory model description	
Task deconstruction and task analysis to develop procedural steps and assessment tools	
Steps of procedures	
Skills and techniques	Course
Supplies, equipment and set-up	
Errors-unambiguously described by quantitative measures	
Pre-test (before skills training begins)	
Technical skills training (psychomotor skills and team training skills)	
Formative assessment with feedback	
Summative assessment with feedback	
Proficiency-based set by repetition until benchmark criteria met-2 consecutive trials	
Debriefing	
Review and learning curve	
Outcomes measures results reporting	

ASSET: Alliance of Surgical Specialties in Education and Training, ACGME: Accreditation Council for Graduate Medical Education

The first two validities (face and content) are determined by a consensus of subject matter experts (SME). The second two validities (concurrent and construct) are determined by comparing the new curriculum to the existing 'gold standard' curriculum, and by comparing whether experts perform better than novices – these comparisons determine not only that the proposed curriculum is as good or better but also whether the curriculum is designed well – if a novice performs as well as an expert on their initial trial, the curriculum's design is poor because there is no value to the training for the novice. Predictive validity is the ultimate goal because it predicts whether the novice was trained well enough to perform in a real clinical situation. It is very difficult to confirm this prediction of clinical outcomes, but by retesting the novice after training to PBP in a similar laboratory skill/task/procedure, on an animal model, can be highly predictive of technical performance in the operating room with a similar procedure. It is noteworthy that although there are other nomenclatures proposed, such as the 2002 American Psychological Association (APA) Standards, nearly all medical literature refers to this long-standing, traditional classification and is familiar to medical researchers and clinicians, and provides the evidence uniquely needed in healthcare, rather than the more recent versions of the APA.

There are a number of other considerations that are important to evaluate even though they are not proof of the quality of the curriculum, but rather reflect how easy it will be to use the curriculum to train/assess novices.

- Reliability – intra-rater reliability - repeated measures produce the same results
- Inter-rater reliability - two or more raters score the same results. Note that an Inter-rater reliability, IRR ≥ 0.80, is accepted as the standard because this corresponds to a 95% confidence level that two raters agree in their evaluation of the curriculum
- Usability – reflects how practical/easy it is to implement or use the course or simulator
- Generalizability – evaluates whether the course can be used in other non-similar situations for which it was originally designed, such as different courses, specialties, etc.

The second aspect of validation involves evaluating whether what is taught in the curriculum is applicable and relevant in the clinical situation. One of the major goals of simulation-based training is to reduce the amount of time a novice will practice and more importantly, reduce or eliminate errors on a patient. By measuring the ToT, it is possible to measure the amount of time in simulation training that is equal to the amount of time training the same skill or task in real life. An important method of representing ToT is the training transfer ratio (TTR), which is expressed as TTR = tr/ts (tr is time of training in the real world and ts is training time in the simulation) for example the TTR in aviation is approximately 0.50, which means that 2 h training time in the simulator is equivalent to 1 h of training in the real aircraft (or ½ or 50% as effective). While this gives a sense of how much cost can be saved by simulation, it does not reflect the extraordinary value of being able to practice very difficult maneuvers or aircraft failures (errors) without the danger of crashing a real aircraft. In healthcare, simulation provides the opportunity to practice difficult skills or procedures without jeopardy to a patient's safety. In addition, it is the manner in which the initial skills (or procedures) can be practiced in a simulated environment which is much less expensive (especially with deliberate practice) than in the actual operating room. By training in simulation, skills can be perfected before attempting procedures on a patient, providing the opportunity to reach proficiency by performing the 'learning curve' (below) outside of the operating room.

A common way of expressing (through visualization) the rate at which a novice learns in PBP simulation is through the learning curve [Figure 1]. This is defined (in Oxford English Dictionary)[11] very concisely as 'a graph showing progress in learning'. By convention, such a graph is represented in Figure 1 - example 1 as a sigmoid curve with the x-axis being the number of trials and the y-axis being the score. Since, the most critical aspect of technical training is performance without error (with error being the 'score'), the graph will have an exponentially decreasing form [Figure 1 - example 2] when graded based upon errors.

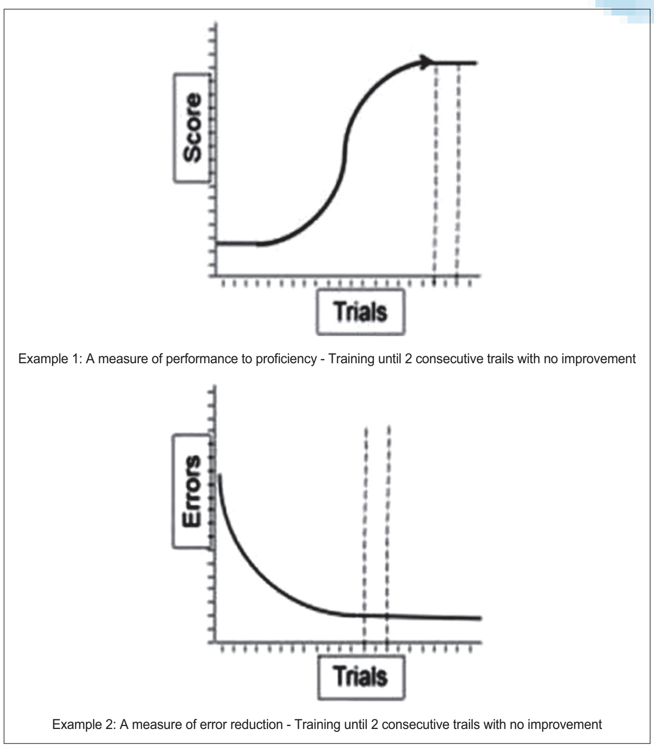

Figure 1: The learning curve

An important part of the training to proficiency is to define what proficiency means and distinguish proficiency from competency. While competent implies an overarching, global level of performance, proficiency implies consistent excellence in a specific area. In the Dreyfus and Dreyfus model of performance[12] in Figure 2, competent in a skill means that the learner can achieve a benchmark level or performance at least once, but not consistently, whereas proficient is determined when a learner can achieve benchmark level consistently, with a minimum of 2 consecutive trials at or above the benchmark level (see below in Expert Benchmark).

There are two important components of the cycle of the full life-cycle development of a curriculum but which are not part of the actual curriculum (course) creation and the training/assessment: They are HST and certification. Their importance lie in the fact, that in addition to actually teaching (knowledge) and training (technical and non-technical skills) to a learner, one of the primary goals is for the learner to be able to pass an independent test demonstrating their competence, and then receive a 'certificate' which is used for numerous reasons, such as on their curriculum vitae, applications for licensure and hospital privileging.

The issue of the HST and the certification are by necessity developed and implemented by (external) organizations other than those responsible for creating a curriculum and training/assessing the learner's performance. In some circumstances, these functions are conducted by a single organization and should that be the case, special precautions are necessary to ensure those who are training the learner are not the same individuals that are performing the HST and/or issuing the certification. The concern about the trainer/assessor performing the HST is that there is an inherent (uncontrollable) bias of the trainer to pass the learner based upon known past performance

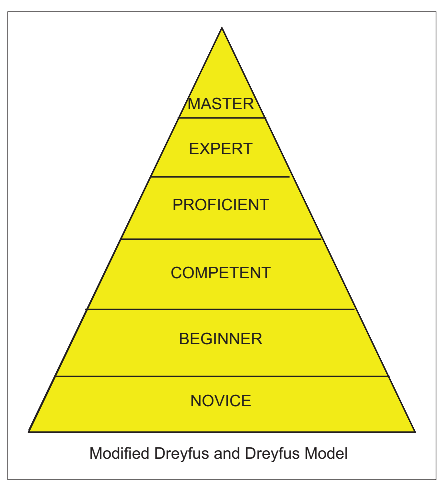

Figure 2: Modified model for Dreyfus and Dreyfus skills acquisition

rather than strictly upon the actual measurement of the HST, especially in the non-technical skills area. For example, following the period of training where the learner demonstrates proficiency to the trainer, during the independent HST the learner's performance does not meet the criteria for proficiency, there is a natural tendency for the trainer, in their role as evaluator of the HST, to take into consideration that the learner had performed to proficiency earlier and therefore pass the learner based on the prior knowledge rather that strictly upon the performance in the HST.

Regarding the issuing of certification, while it is perfectly acceptable for the certifying authority to also be responsible for administering the HST, there is a practical issue of certifying authorities having the resources (financial, personnel, etc.,) to not only develop the HST but also to be able to build and administer a completely independent simulation centre only for the HST. A common alternative (in the United States) is for a professional society in the appropriate specialty (such as Society of American Gastrointestinal and Endoscopic Surgery) to assume the responsibility for the independent development of the HST, and scoring the results, thus providing a verification of knowledge and skills [Table 2] according to the criteria developed by the American College of Surgeons, and which is acceptable to certifying authorities (such as the American Board of Surgery [ABS]) to include in the comprehensive criteria for issuing their (ABS) certification of surgical competence ('Board Certification'). The ABS certificate is the authority which licensing bodies and hospital credentialing/privileging committees permit physicians to practice their surgical specialty.

Table 2: The ACS program for verification of knowledge and skills describes the five levels of verification

Verification of attendance
Verification of satisfactory completion of course objectives
Verification of knowledge and skills
Verification of preceptor experience
Verification of satisfactory patient outcomes

ACS: American College of Surgeons

This seemingly complex system of separate authorities (i.e., training/assessment, HST, certification) is the process of 'checks and balances' by which transparency, abrogation of conflict of interest and evidence of performance can ensure the highest guarantee of an individual learner's competence to practice medicine. In this era of numerous new (and often fraudulent) organizations self-proclaiming and promoting the sale of 'certificates' for profit, it is this process that assures the integrity of the validation of the learner's competence.

METHODS

Technical skills component

The following method describes the fundamental principles and processes by which such a curriculum should be developed. However, implementation of the following principles is complemented by a well-described implementation model entitled 'A stepwise model for simulation-based curriculum development for clinical skills, a modification of the six-step approach',[13] also referred to as the Kern Model. This model is complementary in that it also used the ASSET template.

The process of 'full life-cycle curriculum development' [Table 1] was developed in a consensus conference of the ASSET representatives from 14 surgical specialties, with participation by representatives of the Accreditation Council for Graduate Medical Education and a number of the surgical specialty boards. This template is a distillation of the many decades of simulation development in the numerous high-risk professions mentioned above and the recent experience of simulation in the medical profession. The guiding principles are: (1) The skills must be quantitatively measurable (if possible), or for non-technical skills (teamwork, communication, etc.), the activity to be measured must be unambiguously defined such that two (or more) observers/raters can agree with an IRR ≥ 0.80 and (2) the training and assessment must be to a proficiency benchmark in progressive difficulty, as described under PBP in the introduction. In addition, the curriculum and assessment tools must be validated with the above 5 basic validation types (not necessarily including usability) and must be able to meet requirements of the accreditation authorities.

Needs assessment

The preliminary step for curriculum development begins with the 'needs assessment', as indicated in Table 1. Often the 'need' is obvious – that is, a new surgical technique or device is developed, thus training is necessary. However, to avoid duplication, redundancy, conflict and competition, it is most appropriate to perform an 'inventory' of available literature, publications, etc., to determine whether a curriculum already exists, and whether it meets the 'criteria of the new need' (the requirement). Determining requirement is performed through a consensus conference as to what are the 'ideal' needs and outcomes of the new curriculum. This is compared to the existing curricula (the inventory above), and the analysis is referred to as a 'gap analysis'. If there is a difference between what is desired and what exists (the gap), then an it is clear that a new curriculum must be adapted from existing curriculum, or created from first principles. Obviously, if no current curriculum exists curriculum, then a new curriculum is needed and the process begins as below.

Content of curriculum

The 'content' of the curriculum development process [Figure 3] actually begins with the outcome measures and their metrics. The outcome measures is a description of 'what must be measured' in order to prove the learner is proficient; whereas, the 'metrics' are the actual numbers (or unambiguous definition of the activity). The ideal method of defining the outcome measures is to include input from appropriate authorised governing bodies in the areas of HST, accreditation and credentialing, by inclusion of a representative from an accreditation council (Board), the likelihood that the curriculum will meet official approval is greatly enhanced as a requirement for certification. Clinical SME should be sought through the appropriate surgical society(ies) involved in overseeing all training curriculum. Using a Delphi method[14] to reach consensus on the outcome measures, a consensus must then be reached on which metric (e.g., cm, seconds, defined error, etc.,) to use. As indicated, a numerical ('quantitative') metric is preferable.

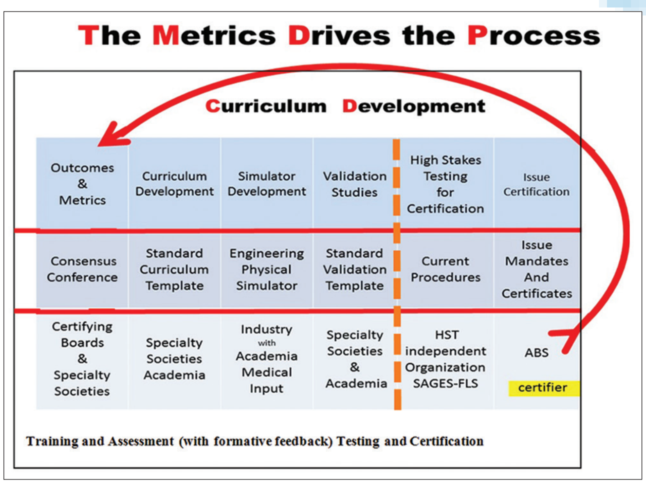

Figure 3: Full life-cycle curriculum development process

The outcome measures/metrics drives the process and the curriculum content including didactic (often on-line) material for the psychomotor skills and (non-technical) TT and C skills. It is essential in the didactic portion of the curriculum to ensure that the 'most common errors' are taught. Again, SMEs derive the best evidence for the didactic portion of the course, as well as, both the psychomotor skills and TT and C skills. Often if a new device or technology is the reason for the curriculum, a new physical or computer-based (including VR) simulator will be developed, which is based upon the outcome measures/metrics. The skills are defined and then grouped into a smaller number of tasks (which are usually composed of more than one skill). This becomes the basis for the simulator course and device development.

Validation study design and trial

Following the final review of the completed course and (if necessary) simulator, a validation study design is performed, often using a consensus conference of SMEs. One of the major results of the design is the 'protocol' which will be used by the individuals or institutes (especially simulation centres) that will be performing the validation trial. The protocol is part of the submission to the Institutional Review Board (IRB), which reviews the quality of the protocol design (usually a double-blinded, randomized controlled trial [RCT] is preferred) and protection (especially the consent form) for the participants (among many other factors). Issues such as recruitment criteria (including inclusion and exclusion criteria) of the participants, randomization and de-identification of the participants' data, database or registry to collect/analyse data, separate curriculum to train the faculty and simulation specialist technicians ('training-the-trainers' course), and organization and administration of multiple institutions are involved and all must be considered. With the approval of the IRB, the validation trial can then begin.

The usual validities (mentioned above) should be used. For face and content validity, which is to determine if the curriculum is going to teach what it was intended to teach, a review should be performed by an external expert panel comprised of SME members who were not involved in the curriculum development. For construct validity, which is to determine if the trial can discriminate between a novice and an expert as well as demonstrating that there is educational value in the curriculum, there should be a comparison of the initial trial of performance from novices to experts. Since the training is to 'proficiency', at the completion of the training/assessment, the novice should have progressed to the level of an experienced surgeon. (Note: If the initial trial of the novices is equivalent to those of the 'experts', then the curriculum is too easy and the novices will not learn by training using this course, since they already perform as well as an 'expert'. Often a pilot study performed with a few subjects in each group (novice/expert) to estimate construct validity can demonstrate the likelihood that there is construct validity). For concurrent validity, which demonstrates whether the curriculum is as good as (or better than) the current standard of practice, there should be a comparison of the performance of a control group with the novice group. The control group uses whatever is the current curriculum (if any) and the novice group uses the new curriculum. With predictive validity, which is intended to show that based upon the novice performance to proficiency, it is possible to predict their performance in a similar clinical situation. Predictive validity is rather difficult to acquire because it requires that the performance of the novice (after reaching proficiency) will be as good as an experienced surgeon based upon clinical outcomes. This can be in an animal model, which could prove that technical performance on a simulator will predict the learners' technical performance in an animal laboratory setting. However, for clinical situations, the follow-up must be much longer (with many more 'trials' conducted on real patients) because the 'outcome measures' are based on factors such as morbidity and mortality.

Two steps proceed the psychomotor skills portion – the expert benchmark and the pre-test.

Expert benchmark setting

The experts in the trial set the benchmark (a value which the novices must achieve on each task before progressing to the next task) this is the meaning and definition of 'proficiency-based progression'. This is determined by the learning curve of the expert (experienced surgeon) for each of the tasks of the validation trial [Figure 4]. Using the Dreyfus and Dreyfus model, of levels of proficiency, the mean value of all the experts is defined as the of levels of proficiency, level. Below proficiency, is competent and beginner whereas above proficient is a true expert and master. The 'standard deviations' are notional, and are used to speculate that it may be possible to quantify the Dreyfus & Dreyfus performance levels, though it is likely to be more complex than simple standard deviations as depicted here. The merit and value to this model is that it would be a valuable objective measurement tool during the training and assessment of a learner to estimate to which level in their learning curve of their performance they have progressed. (Note: The original Dreyfus and Dreyfus model did not include master; however, subsequent publications did include this. The discrimination between the two is that experts are not only consistent and proficient in the skill/task, but can recognize and perform variations successfully; whereas Masters are capable of inventing wholly new adaptations or complete techniques). This benchmark value (whatever the metric) must be achieved (i.e., the score for the learner is 100% of the benchmark value). The benchmark values should include both correct performance and absence of (or minimal non-critical) errors. It is essential to define which errors are 'critical', for a single critical error results automatically in to failure.

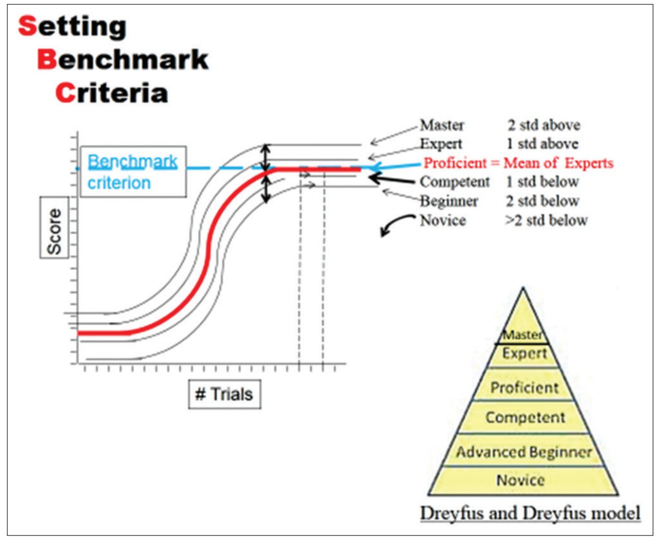

Figure 4: Expert benchmark criteria

Pre-test and post-test

The beginning of the trial is with a pre-test for both control and experimental (new curriculum) groups. This is essential in order to establish a baseline of performance for each learner. This test should be different than the tasks which are to be trained/assessed, but similar enough. In the use of simulation, the pre-test is frequently an animal tissue model (often from a butcher's shop). The purpose is to determine if the learner (in both control and experimental groups) perform better on this same test model after the training period, which demonstrates that they have improved their performance, and is a direct measure of the educational value of the curriculum. Comparison of the control group to the new curriculum group demonstrates concurrent validity – whether and how much better the new curriculum is compared to the current standard of training.

Training to proficiency

The two components of training to proficiency are deliberate[9] practice until the benchmark is achieved and formative feedback. Whenever an error is made during the training, the learner is immediately informed of the error and how to avoid such an error in the next trial. Thus, each learner performs each task (with formative feedback until proficiency is reached), progressing from the initial simple tasks to the more difficult final tasks. It is after completion of all the simulated tasks that the post-test is performed. Since the learner has just demonstrated that he/she can perform each task proficiently on the simulator, they should be able to perform the post-test with maximum performance and minimal to no errors.

Non-technical skills component

The complexity of today's procedures is such that a team approach is necessary, and this requires training in non-technical skills - both teamwork and communication skills. Other aspects are often included as well, such as leadership, professionalism, etc., however, this emphasis is upon TT and C. For this, there is a long history of successful training through the use of scenarios and often performed with manikins. The use of scenarios revolves around a difficult task/procedure, or crisis management (such as cardiac arrest, anaesthetic complication, converting the procedure to open surgery, etc.). The training involves multiple professionals on the team – physicians, nurses, technicians, etc., from different specialties (anaesthesia, surgery, nursing, etc.). Currently, the emphasis is upon 'inter-professional training', where the members of the team that usually work together are trained together at the same time with the various scenarios. A number of scenarios have been created, as well as different approaches; however, the TeamSTEPPS method is well accepted and can be adapted to PBP training and assessment.

The standard today is to train the team on a 'scenario', usually with videotaping, and upon completion of the scenario, a debriefing of the team performance is conducted, using the videotape review with the learners as well as (occasionally) checklists of specific tasks, activities and errors. There is great merit in this approach, and the students do learn; however, there is no quantitative assessment – at the completion of the debriefing the team is finished, and there is no determination of pass or fail – this is a very valuable educational experience but not as training for proficiency. Current unpublished efforts are underway to develop a PBP approach to TT and C.

CONCLUSION

The introduction of simulation for procedural skills training has permitted quantitative training and evaluation of a learner's skill performance. Based upon decades of experience in the non-healthcare field with simulation, the method is presented as a candidate for the next generation of procedural skills training. While a refinement is expected, the basic principles behind the PBP approach are sound and have been proven in other industries. Application of this process, using the measurement of performance to proficiency, coupled with the avoidance of errors, will result in an improvement in patient safety.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Gallagher AG, O'Sullivan GC. Fundamentals of Surgical Simulation: Principles and Practice (Improving Medical Outcomes – Zero Tolerance). London: Springer Verlag; 2012.
2. Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G, *et al*. Virtual reality simulation for the operating room: Proficiency-based training as a paradigm shift in surgical skills training. Ann Surg 2005;241:364-72.
3. Gaba DM, DeAnda A. A comprehensive anesthesia simulation environment: Re-creating the operating room for research and training. Anesthesiology 1988;69:387-94.
4. TeamSTEPPS: National Implementation. Available from: http://www.teamstepps.ahrq.gov/. [Last accessed on 2015 Jun 18].
5. Satava RM. Virtual reality surgical simulator. The first steps. Surg Endosc 1993;7:203-5.
6. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. Am J Surg 1997;173:226-30.
7. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, *et al*. Virtual reality training improves operating room performance: Results of a randomized, double-blinded study. Ann Surg 2002;236:458-63.
8. Fundamentals of Laparoscopic Surgery. Available from: http://www.flsprogram.org. [Last accessed on 2015 Sep 25].
9. Ericsson KA. Development of Professional Expertise: Toward Measurement of Expert Performanc and Design of Optimal Learning Cambridge University Press: New York 2009.
10. Zevin B, Levy JS, Satava RM, Grantcharov TP. A consensus-based framework for design, validation, and implementation of simulation-based training curricula in surgery. J Am Coll Surg 2012;215:580-6.e3.
11. Oxford English Dictionary. Available from: http://www.oed.com. [Last accessed on 2015 Sep 25].
12. Dreyfus SE, Dreyfus HL. A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition. Washington, DC: Storming Media; 1980.
13. Khamis NN, Satava RM, Alnassar SA, Kern DE. A stepwise model for simulation-based curriculum development for clinical skills, a modification of the six-step approach. Surg Endosc 2015. [Epub ahead of print].
14. Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. J Clin Epidemiol 2003;56:1150-6.